

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:28 ; Search time 85.6067 Seconds

(without alignments)
3062.893 Million cell updates/sec

Title: US-10-657-280-1

Perfect score: 4907

Sequence: 1 MFLKLRQRYTCLSHRYGL.....SLKQLAVFTDSKNYGRQLK 928

Scoring table:

BLOSUM62

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : A_Geneseq_29Jan04:*

1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4907	100.0	928	4	ABR61376
2	4907	100.0	928	7	ADD27812
3	4907	100.0	928	7	ABW01488
4	4907	100.0	928	7	ABW01537
5	4896	99.8	1256	5	AAE25290
6	4630	94.4	1199	7	ADD27810
7	4630	94.4	1199	7	ABW01487
8	4630	94.4	1199	7	ABW01536
9	4500	91.7	1196	5	AAE25294
10	4468	91.1	846	6	ADA54824
11	4391	89.5	1459	6	ABU07381
12	3894	79.4	908	4	ABR61384
13	3894	79.4	908	7	ADD27817
14	3894	79.4	908	7	ABW01491
15	3894	79.4	908	7	ABW01540
16	1900	38.7	367	4	AAH16457
17	1900	38.7	367	4	AAH16457
18	1900	38.7	367	4	AAH16457
19	1900	38.7	367	4	AAH16457
20	1900	38.7	367	4	AAH16457
21	1900	38.7	367	4	AAH16457
22	1286	26.2	242	7	ADD26807
23	923	18.8	178	6	ABU11844
24	503	10.3	113	4	ABR61382
25	503	10.3	113	7	ADD27822

26	503	10.3	113	7	ABW01494	Abw01494	Partial r
27	503	10.3	113	7	ABW01543	Abw01543	Partial r
28	488	9.9	652	4	ABR59094	AbB59094	Drosophila
29	479	9.8	502	4	ABR61383	AbR61383	Partial D
30	479	9.8	502	7	ADD27824	Add27824	Fruit fly
31	479	9.8	502	7	ABW01495	Abw01495	Partial f
32	479	9.8	502	7	ABW01544	Abw01544	Partial f
33	349	7.1	68	4	ABR31817	AbB31817	peptide #
34	349	7.1	68	4	AAH16416	AAH16416	peptide #
35	349	7.1	68	4	AAH17137	AAH17137	Human bon
36	349	7.1	68	4	AAH58821	AAH58821	Human bra
37	349	7.1	68	4	ABG53045	ABG53045	Human liv
38	349	7.1	68	5	ABG41140	ABG41140	Human pep
39	265	5.4	545	2	AAW79296	AAW79296	Neisseria
40	255.5	5.2	546	7	ADD74499	AdB74499	Mycobacte
41	251	5.1	63	6	ABR63220	ABR63220	Cione T11
42	251	5.1	63	6	ABR43797	ABr43797	Cione T11
43	240	4.9	364	6	ADA34095	Ada34095	Acinetoba
44	173.5	3.5	2469	5	AAE18207	Aae18207	Human MOL
45	173.5	3.5	2469	7	ADD18192	Add18192	Human mol

ALIGNMENTS

RESULT 1	ABR61376	standard; protein; 928 AA.
ID	ABR61376;	
AC	ABR61376;	
XX		
DT	01-AUG-2003	(first entry)
XX		
DE	Human GlcNAc-6-phosphotransferase alpha-subunit.	
XX		
KW	Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;	
KW	GlcNAc-6-phosphotransferase; phosphodiester alpha-GlcNAcase;	
KW	N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;	
KW	enzyme replacement therapy; phosphorylated lysosomal hydrolase;	
KW	lysosomal storage disease; enzyme; alpha-subunit.	
XX		
OS	Homo sapiens.	
XX		
PN	US6537785-B1.	
XX		
PD	25-MAR-2003.	
XX		
PF	10-AUG-2000; 2000US-00636077.	
XX		
PR	14-SEP-1999; 99US-0153831P.	
XX		
PA	(GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.	
XX		
XX	Canfield WM;	
PI	WPI; 2001-290356/30.	
XX		
DR	N-PSDB; ACC61001.	
XX		
PT	Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.	
PT		
PS	Disclosure; Page 26-28; 62pp; English.	
XX		
CC	The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-6-phosphotransferase) (I) and phosphodiester alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (II). The protein of the invention has nephrotropic activity, and may be useful in enzyme replacement therapy. A protein of the invention (I), (II) is useful for preparing a phosphorylated lysosomal hydrolase. The phosphorylated hydrolase comprising a terminal mannose-6-phosphate, is useful for treating a	

CC patient suffering from a lysosomal storage disease. The present sequence
 CC is used in the exemplification of the invention

XX Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 4; Length 928;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MFLKIQRTYTCLSHRVGLVYCFGLGVVTVTSARQFGVVLWMSRDQYHVLFDSDYRNI 60
DB 1 MFLKIQRTYTCLSHRVGLVYCFGLGVVTVTSARQFGVVLWMSRDQYHVLFDSDYRNI 60
QY 61 AGKSFQNRCLPMPIDVYTWNGTDLLELKELOQVREMEEOQAMEIIGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYTWNGTDLLELKELOQVREMEEOQAMEIIGKNTTEPTK 120
QY 121 KSEKQLECLTHCIKVPMLVLDPALPANITLKDVPSLYPSFASADIFNVAKPKPSTNV 180
DB 121 KSEKQLECLTHCIKVPMLVLDPALPANITLKDVPSLYPSFASADIFNVAKPKPSTNV 180
QY 181 SYVVPSTKDVEDASGLKNSROTWVRGYLTTEKEVPGVLVMDLAFLSGFPPTFKET 240
DB 181 SYVVPSTKDVEDASGLKNSROTWVRGYLTTEKEVPGVLVMDLAFLSGFPPTFKET 240
QY 241 NQKTKLPENLSKVKLLQLYSEASVALIKLNPPDFOELNKQTKNMTIDGKELTISPA 300
DB 241 NQKTKLPENLSKVKLLQLYSEASVALIKLNPPDFOELNKQTKNMTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKQEDISASRFEDNEELRYSLSIERHAPVNRNIFVTNGQIPSWNL 360
DB 301 YLLMDLSAISQSKQEDISASRFEDNEELRYSLSIERHAPVNRNIFVTNGQIPSWNL 360
QY 361 DNPRTVITHQDFENLSHLPTFSSPAIESHIRELEGSOKFIYLDNDVMEGXVWPDF 420
DB 361 DNPRTVITHQDFENLSHLPTFSSPAIESHIRELEGSOKFIYLDNDVMEGXVWPDF 420
QY 421 YHSKQGVYLTWPVPCNAGCGPMSIKDGYCDKACNNACDMDGDCSGNSGGRYIAG 480
DB 421 YHSKQGVYLTWPVPCNAGCGPMSIKDGYCDKACNNACDMDGDCSGNSGGRYIAG 480
QY 481 GGGTSGIGVGHPMQGGGINSVSYCNOGCANSMWLDKFCDDQACNVLSGCFPAGDCGQDH 540
DB 481 GGGTSGIGVGHPMQGGGINSVSYCNOGCANSMWLDKFCDDQACNVLSGCFPAGDCGQDH 540
QY 541 HELYKVIILPNOTHYIIPKGECLPYFSPAFAVAKGVEGAYSDNPIIRHASJANKWKTIHL 600
DB 541 HELYKVIILPNOTHYIIPKGECLPYFSPAFAVAKGVEGAYSDNPIIRHASJANKWKTIHL 600
QY 601 IMHSGNATTIHENLTFQNTNDEEFMOTIYEVDRREGPKLNSTQKGYENLVSPITLP 660
DB 601 IMHSGNATTIHENLTFQNTNDEEFMOTIYEVDRREGPKLNSTQKGYENLVSPITLP 660
QY 661 EAEILFEDIPEKERPFKRDVNSTRAQEBVKIPLVNISLPPDAQLSANTLDLOEH 720
DB 661 EAEILFEDIPEKERPFKRDVNSTRAQEBVKIPLVNISLPPDAQLSANTLDLOEH 720
QY 721 GPITLKGYNLSKALRSFLMNSQAKIKNOAIITDENRDSIVAQOEQVHKSLIPNSLG 780
DB 721 GPITLKGYNLSKALRSFLMNSQAKIKNOAIITDENRDSIVAQOEQVHKSLIPNSLG 780
QY 781 VSERLQRLTFPAVSVKNGHOGONPDLLETTAFRVEHTQKTIGAVTEKPEPSLIIV 840
DB 781 VSERLQRLTFPAVSVKNGHOGONPDLLETTAFRVEHTQKTIGAVTEKPEPSLIIV 840
QY 841 PLESQVTKKKITGKEKENSMEENAEHIGTBYLGRKIOHTYDSVLGLPWEKKKYF 900
DB 841 PLESQVTKKKITGKEKENSMEENAEHIGTBYLGRKIOHTYDSVLGLPWEKKKYF 900
QY 901 ODLDEESLKTOLAVFTDSKKTGRQK 928
DB 901 ODLDEESLKTOLAVFTDSKKTGRQK 928
QY 901 QDLDBEBSLKTOLAVFTDSKKTGRQK 928
DB 901 QDLDBEBSLKTOLAVFTDSKKTGRQK 928

```

RESULT 2

ID ADD27812 standard; protein; 928 AA.

AC ADD27812;

DT 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase alpha subunit.

KM human; protein phosphorylation; soluble GlcNAc-phosphotransferase;

XM UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.

OS Homo sapiens.

PN US2003119088-A1.

PD 26-JUN-2003.

PF 21-DEC-2001; 2001US-00023888.

PR 21-DEC-2001; 2001US-00023888.

PA (NOVA-) NOVAZYME PHARM INC.

PI Canfield W, Kudo M;

DR N-PSDB; ADD27811.

PT Phosphorylating a protein for treating a patient suffering from a

PT lysosomal storage disease e.g. Fabry's disease by contacting the protein

PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated

PT protein.

PS Claim 7; SEQ ID NO 4; 55pp; English.

XX The invention relates to a method of phosphorylating a protein comprising

XX contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-

XX acetylglucosamine) and producing a phosphorylated protein. The method is

XX useful for treating a patient suffering from a lysosomal storage disease

XX e.g. Fabry's disease. The present sequence represents the amino acid

XX sequence of the human GlcNAc-phosphotransferase alpha subunit.

XX Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 7; Length 928;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MFLKIQRTYTCLSHRVGLVYCFGLGVVTVTSARQFGVVLWMSRDQYHVLFDSDYRNI 60
DB 1 MFLKIQRTYTCLSHRVGLVYCFGLGVVTVTSARQFGVVLWMSRDQYHVLFDSDYRNI 60
QY 61 AGKSFQNRCLPMPIDVYTWNGTDLLELKELOQVREMEEOQAMEIIGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYTWNGTDLLELKELOQVREMEEOQAMEIIGKNTTEPTK 120
QY 121 KSEKQLECLTHCIKVPMLVLDPALPANITLKDVPSLYPSFASADIFNVAKPKPSTNV 180
DB 121 KSEKQLECLTHCIKVPMLVLDPALPANITLKDVPSLYPSFASADIFNVAKPKPSTNV 180
QY 181 SYVVPSTKDVEDASGLKNSROTWVRGYLTTEKEVPGVLVMDLAFLSGFPPTFKET 240
DB 181 SYVVPSTKDVEDASGLKNSROTWVRGYLTTEKEVPGVLVMDLAFLSGFPPTFKET 240
QY 241 NQKTKLPENLSKVKLLQLYSEASVALIKLNPPDFOELNKQTKNMTIDGKELTISPA 300
DB 241 NQKTKLPENLSKVKLLQLYSEASVALIKLNPPDFOELNKQTKNMTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKQEDISASRFEDNEELRYSLSIERHAPVNRNIFVTNGQIPSWNL 360
DB 301 YLLMDLSAISQSKQEDISASRFEDNEELRYSLSIERHAPVNRNIFVTNGQIPSWNL 360

```

```
DB 301 YLIMDLAISQSKODEDISASRPEDNEELRYSLSIRHAPWVRNI FIVTNGQIPSWMLN 360
QY 361 DNRVYIVTHQDVFRNLISHLPTSSPAIESHIRIEGLSQKFIYNDVMPFGDVPDPDF 420
DB 361 DNRVYIVTHQDVFRNLISHLPTSSPAIESHIRIEGLSQKFIYNDVMPFGDVPDPDF 420
QY 421 YSHSKQKXYLLTPVPCNCAEGCGSWIKDGYCDKACNNSACMDGDCGSGSGSRYIAG 480
DB 421 YSHSKQKXYLLTPVPCNCAEGCGSWIKDGYCDKACNNSACMDGDCGSGSGSRYIAG 480
QY 481 GGGTSGIGVGHMPQFGGINSVSYCNQGCANSMWLADKFCDOACNVLSGCFDAGDCGQDHF 540
DB 481 GGGTSGIGVGHMPQFGGINSVSYCNQGCANSMWLADKFCDOACNVLSGCFDAGDCGQDHF 540
QY 541 HELYKVIILPNQTHYIIPKGECLPYPSFAVAKRGVEGAYSNDPIIRHASIANKKXTIHL 600
DB 541 HELYKVIILPNQTHYIIPKGECLPYPSFAVAKRGVEGAYSNDPIIRHASIANKKXTIHL 600
QY 601 IMHSGMNAITIHFNLTFCQNTNDEEFPMQITVEVDTRREGPKLNSTQKGYENLVSPITLLP 660
DB 601 IMHSGMNAITIHFNLTFCQNTNDEEFPMQITVEVDTRREGPKLNSTQKGYENLVSPITLLP 660
QY 661 EAEILFEDIPEKRFPEKFRHDVNSTRAQOEVKIPLVNIISLPKDAQLSINTLDPQLEH 720
DB 661 EAEILFEDIPEKRFPEKFRHDVNSTRAQOEVKIPLVNIISLPKDAQLSINTLDPQLEH 720
QY 721 GGIITLKGYNLSKALIRSFILNMSQAKIKNOAIITDETNDLSIAPQOEKQVHKSIIPLNSLG 780
DB 721 GGIITLKGYNLSKALIRSFILNMSQAKIKNOAIITDETNDLSIAPQOEKQVHKSIIPLNSLG 780
QY 781 VSEIRLQRLTFPAVSVKVNGHDQGNPDLLETTAFRVEYTHQKTIIGNVTKERKPSLIIV 840
DB 781 VSEIRLQRLTFPAVSVKVNGHDQGNPDLLETTAFRVEYTHQKTIIGNVTKERKPSLIIV 840
QY 841 PLESQVTKERKTIIGKEKENSMEENAENHIGYTEVLGRKLIQHYTDSYLGFLPWEKKCYF 900
DB 841 PLESQVTKERKTIIGKEKENSMEENAENHIGYTEVLGRKLIQHYTDSYLGFLPWEKKCYF 900
QY 901 ODLLDEESLKTQOLAVFTDSKNTGRQK 928
DB 901 ODLLDEESLKTQOLAVFTDSKNTGRQK 928

RESULT 3
ABW01488
ID ABW01488 standard; protein; 928 AA.
XX
AC ABW01488;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human GlcNAc-phosphotransferase alpha subunit precursor protein.
XX
KW Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
KW galectin; human; enzyme; lectin resistant cell;
KW deoxymannojirimycin; kifunensine; glycosylation inhibition.
OS
XX Homo sapiens.
XX
PN US2003124652-A1.
XX
PD 03-JUL-2003.
XX
PF 21-DEC-2001; 2001US-00023889.
XX
PR 21-DEC-2001; 2001US-00023889.
XX
PA (NOVA-) NOVARTIS PHARM INC.
XX
PI Canfield MW;
XX
DR WPI; 2003-810984/76.
```

```
DR N-PSDB; AAD62491.
XX
PT Producing a high mannose glycoprotein for treating lysosomal storage
PT disease, comprises culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
XX
PS Claim 10; Page 15-18; 46pp; English.
XX
CC The invention relates to a method for producing a high mannose
CC glycoprotein. The method comprises: introducing and expressing a
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
CC the cell in the presence of a lectin to obtain a lectin resistant cell;
CC isolating the cell; culturing the cell in the presence of
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
CC glycoprotein; and collecting the glycoprotein. The invention is useful
CC in gene therapy. The method is useful for producing a high mannose
CC glycoprotein in a complex carbohydrate deficient cell for treating
CC lysosomal storage disease. The present sequence is human N-
CC acetylglucosamine-1 (GlcNAc)-phosphotransferase alpha subunit precursor
CC protein
XX
SQ Sequence 928 AA;
Query Match 100.0%; Score 4907; DB 7; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLFLLQROTYYTCLSHRYGLVYCFLGVVYIVSAFQGEVWLEMSRDQYVHLPDSYRDN 60
DB 1 MLFLLQROTYYTCLSHRYGLVYCFLGVVYIVSAFQGEVWLEMSRDQYVHLPDSYRDN 60
QY 61 AGKSFQNRRLCPMPIDVYTVVNGTDLLELKLQOVRBQMBEQKAREILGNTEPTK 120
DB 61 AGKSFQNRRLCPMPIDVYTVVNGTDLLELKLQOVRBQMBEQKAREILGNTEPTK 120
QY 121 KSEKQECCLTHGCKVPMCLVLDPALPANITLKVPSYPSFHSASIPFNKPNSTNV 180
DB 121 KSEKQECCLTHGCKVPMCLVLDPALPANITLKVPSYPSFHSASIPFNKPNSTNV 180
QY 181 SVVVPDSTKQVEDAHSGLLKGNRSQTVWAGYLTDDXEPGLVIMQDLAFLSGFPPTKET 240
DB 181 SVVVPDSTKQVEDAHSGLLKGNRSQTVWAGYLTDDXEPGLVIMQDLAFLSGFPPTKET 240
QY 241 NOLKTKLPEULSSKVKLLQYSSASVALLKNNPKDPOELNOKTKNMITDGEILTISA 300
DB 241 NOLKTKLPEULSSKVKLLQYSSASVALLKNNPKDPOELNOKTKNMITDGEILTISA 300
QY 301 YLLMDLSAISQSKODEDISASRPEDNEELRYSLSIRHAPWVRNI FIVTNGQIPSWMLN 360
DB 301 YLLMDLSAISQSKODEDISASRPEDNEELRYSLSIRHAPWVRNI FIVTNGQIPSWMLN 360
QY 361 DNRVYIVTHQDVFRNLISHLPTSSPAIESHIRIEGLSQKFIYNDVMPFGDVPDPDF 420
DB 361 DNRVYIVTHQDVFRNLISHLPTSSPAIESHIRIEGLSQKFIYNDVMPFGDVPDPDF 420
QY 421 YSHSKQKXYLLTPVPCNCAEGCGSWIKDGYCDKACNNSACMDGDCGSGSGSRYIAG 480
DB 421 YSHSKQKXYLLTPVPCNCAEGCGSWIKDGYCDKACNNSACMDGDCGSGSGSRYIAG 480
QY 481 GGGTSGIGVGHMPQFGGINSVSYCNQGCANSMWLADKFCDOACNVLSGCFDAGDCGQDHF 540
DB 481 GGGTSGIGVGHMPQFGGINSVSYCNQGCANSMWLADKFCDOACNVLSGCFDAGDCGQDHF 540
QY 541 HELYKVIILPNQTHYIIPKGECLPYPSFAVAKRGVEGAYSNDPIIRHASIANKKXTIHL 600
DB 541 HELYKVIILPNQTHYIIPKGECLPYPSFAVAKRGVEGAYSNDPIIRHASIANKKXTIHL 600
QY 601 IMHSGMNAITIHFNLTFCQNTNDEEFPMQITVEVDTRREGPKLNSTQKGYENLVSPITLLP 660
DB 601 IMHSGMNAITIHFNLTFCQNTNDEEFPMQITVEVDTRREGPKLNSTQKGYENLVSPITLLP 660
QY 661 EAEILFEDIPEKRFPEKFRHDVNSTRAQOEVKIPLVNIISLPKDAQLSINTLDPQLEH 720
```

Db 661 EAEILFEDIPKEKRPKFRKHDVNSTRAOEKVIPLVNIISLPDQALSLNTLDQLQEH 720
 QY 721 GDTLLKGYVLSKSLALRSFLMNSQAKIKNOAIITDETNDLSLVAPQEQVHKSLIIPNSLG 780
 Db 721 GDTLLKGYVLSKSLALRSFLMNSQAKIKNOAIITDETNDLSLVAPQEQVHKSLIIPNSLG 780
 QY 781 VSERLQRLTFPAVSVKVNQHDGQNPPLDLETTARFRVETHQKTIIGAVTKEKPPSLIV 840
 Db 781 VSERLQRLTFPAVSVKVNQHDGQNPPLDLETTARFRVETHQKTIIGAVTKEKPPSLIV 840
 QY 841 PLESQMTKEKKITGKEKENSMEENAHNIGVTEVLGRKLOHYTDSYLGFLPMEKKCYF 900
 Db 841 PLESQMTKEKKITGKEKENSMEENAHNIGVTEVLGRKLOHYTDSYLGFLPMEKKCYF 900
 QY 901 ODLLDEESLKTOLAYFTDSKNTGRQLK 928
 Db 901 ODLLDEESLKTOLAYFTDSKNTGRQLK 928

RESULT 4

ABW01537 standard; protein; 928 AA.
 ID ABW01537
 AC ABW01537;
 DT 15-JUN-2004 (first entry)
 DE Human GlcNAc-phosphotransferase alpha subunit precursor protein.
 KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
 KM N-acetylglucosamine-1-phosphotransferase; enzyme; gene therapy; human.
 OS Homo sapiens.
 PN US2003124653-A1.
 PD 03-JUL-2003.
 PF 21-DEC-2001; 2001US-00023890.
 PR 21-DEC-2001; 2001US-00023890.
 PA (NOVA-) NOVAZYME PHARM INC.
 PI Canfield WM;
 DR WPI, 2003-810985/76.
 DR N-PDB; AAD62650.
 PT Producing a glycoprotein with reduced complex carbohydrates by culturing
 PT the lectin resistant mammalian cell expressing the glycoprotein for
 PT treating lysosomal storage disease.
 PS Claim 10; Page 16-18; 46pp; English.
 CC The present invention provides a method of producing a glycoprotein
 CC having reduced complex carbohydrates by culturing the lectin resistant
 CC mammalian cell expressing the glycoprotein. The method is useful for
 CC producing a glycoprotein with reduced complex carbohydrates for treating
 CC lysosomal storage disease. The present invention is also useful in gene
 CC therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-
 CC phosphotransferase alpha subunit precursor protein
 CC
 SQ Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 7; Length 928;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLEKLLQRTYTCLSHRVGLVYVCFGVVTVIVSAFQFGSEVLEMSRDQYHVLFDSDYDNI 60
 Db 1 MLEKLLQRTYTCLSHRVGLVYVCFGVVTVIVSAFQFGSEVLEMSRDQYHVLFDSDYDNI 60

QY 61 AGSFQNRCLPMPIDVYVTVWNGTDLLELLELQVREOMEEQKAMREILCKNTTEPTK 120
 Db 61 AGSFQNRCLPMPIDVYVTVWNGTDLLELLELQVREOMEEQKAMREILCKNTTEPTK 120
 QY 121 KSEKQLECLTHQIKYPMILVLPALPANITLKDVSGLVSPFASQDIPVAKPKPSTNV 180
 Db 121 KSEKQLECLTHQIKYPMILVLPALPANITLKDVSGLVSPFASQDIPVAKPKPSTNV 180
 QY 181 SVVPFDSTKDVEDAHSGLLKGNRSQRTVMRGYLTIDKEVGLVIMODLAFSLGFPPEKST 240
 Db 181 SVVPFDSTKDVEDAHSGLLKGNRSQRTVMRGYLTIDKEVGLVIMODLAFSLGFPPEKST 240
 QY 241 NQKTLPLPNSKXVLLQYSEASVALKLNPNPDQGLNKOTKKNMTIDKELITISA 300
 Db 241 NQKTLPLPNSKXVLLQYSEASVALKLNPNPDQGLNKOTKKNMTIDKELITISA 300
 QY 301 YLLMDLSAISQSKODEDISASRPEDNEELRYSLSIERHAPVWRNIFIVTNGQIPSWNL 360
 Db 301 YLLMDLSAISQSKODEDISASRPEDNEELRYSLSIERHAPVWRNIFIVTNGQIPSWNL 360
 QY 361 DNEPRTIIVTHQVFNRLSHLPTSSPALSHIRIEGLSQKFIYANDVMFGKVPDPDF 420
 Db 361 DNEPRTIIVTHQVFNRLSHLPTSSPALSHIRIEGLSQKFIYANDVMFGKVPDPDF 420
 QY 421 YSHSKQKQYTLTPVPCABEGCGSWIKDGYCDKACNNSACMDGDGCGNSGGSRYIAG 480
 Db 421 YSHSKQKQYTLTPVPCABEGCGSWIKDGYCDKACNNSACMDGDGCGNSGGSRYIAG 480
 QY 481 GGGTGSIGVGHWPQFGGINSVYCNQGCANSWLADKFCDOACNVLSGCFDAGCGQDHF 540
 Db 481 GGGTGSIGVGHWPQFGGINSVYCNQGCANSWLADKFCDOACNVLSGCFDAGCGQDHF 540
 QY 541 HELYKYLILPNQTHYIIPRGECLPYSPFAVAKRGVGSYSDNPRIIRHASICNKKYTHL 600
 Db 541 HELYKYLILPNQTHYIIPRGECLPYSPFAVAKRGVGSYSDNPRIIRHASICNKKYTHL 600
 QY 601 IMHSGNATTIHNLTFQNTNDEEFKQITVEVDTRBEGKLNSTAKQYENLVSPITLLP 660
 Db 601 IMHSGNATTIHNLTFQNTNDEEFKQITVEVDTRBEGKLNSTAKQYENLVSPITLLP 660
 QY 661 EAEILFEDIPKEKRPKFRKHDVNSTRAOEKVIPLVNIISLPDQALSLNTLDQLQEH 720
 Db 661 EAEILFEDIPKEKRPKFRKHDVNSTRAOEKVIPLVNIISLPDQALSLNTLDQLQEH 720
 QY 721 GDTLLKGYVLSKSLALRSFLMNSQAKIKNOAIITDETNDLSLVAPQEQVHKSLIIPNSLG 780
 Db 721 GDTLLKGYVLSKSLALRSFLMNSQAKIKNOAIITDETNDLSLVAPQEQVHKSLIIPNSLG 780
 QY 781 VSERLQRLTFPAVSVKVNQHDGQNPPLDLETTARFRVETHQKTIIGAVTKEKPPSLIV 840
 Db 781 VSERLQRLTFPAVSVKVNQHDGQNPPLDLETTARFRVETHQKTIIGAVTKEKPPSLIV 840
 QY 841 PLESQMTKEKKITGKEKENSMEENAHNIGVTEVLGRKLOHYTDSYLGFLPMEKKCYF 900
 Db 841 PLESQMTKEKKITGKEKENSMEENAHNIGVTEVLGRKLOHYTDSYLGFLPMEKKCYF 900
 QY 901 ODLLDEESLKTOLAYFTDSKNTGRQLK 928
 Db 901 ODLLDEESLKTOLAYFTDSKNTGRQLK 928

RESULT 5

AAE25290 standard; protein; 1256 AA.

AAE25290;

30-OCT-2002 (first entry)

QY Human nucleic acid-associated protein (NAAP-9).
 DE Human, nucleic acid-associated protein; NAAP-9; neurological disorder;
 KM arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;

KM Lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
 KM autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;
 KM gene therapy; nootropic; neuroprotective; cerebroprotective; virucide;
 KM immunosuppressive; protozoacide; antimicrobial.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Domain 448..469
 FT /note="Notch domain"
 FT Domain 500..536
 FT /note="Notch domain"
 FT Domain 1018..1030
 FT /note="EF-hand calcium-binding domain"
 PN MO200250279-A2.
 XX
 PD 27-JUN-2002.
 XX
 PD 19-DEC-2001; 2001WO-US050256.
 XX
 PF 21-DEC-2000; 2000US-0257714P.
 XX
 PR 05-JAN-2001; 2001US-0260081P.
 PR 16-JAN-2001; 2001US-026302P.
 PR 23-JAN-2001; 2001US-0263823P.
 PR 02-FEB-2001; 2001US-0266088P.
 PR 29-OCT-2001; 2001US-0348442P.
 XX
 PA (INCYTE GENOMICS INC.
 XX
 PI Baughin MR, Lu Y, Arvizu C, Rankumar J, Yao MG, Policky JL;
 PI Walla NK, Tribolety KM, Yue H, Batra S, Ding L, Lal PG;
 PI Borowsky JM, Lu DM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;
 PI Glezen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalila AJA;
 PI Lee EA, Yang J, Gorvad AE, Emerling BM, Marquis JP, Lee SY;
 PI Swarnakar A, Reddy R;
 XX
 DR WPI; 2002-519887/55.
 DR N-PSDB; AAD41199.
 XX
 PT Nucleic acid associated proteins and nucleic acids for diagnosing,
 PT treating and preventing cell proliferative (e.g. cancers), neurological
 PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).
 PS Claim 64; Page 163-165; 193pp; English.
 XX
 CC The invention relates to nucleic acid-associated proteins (NAAP) and
 CC nucleic acids. The nucleic acid and amino acid sequences are useful for
 CC diagnosing, treating and preventing cell proliferative e.g.
 CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological
 CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and
 CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections
 CC (e.g. malaria, or leishmania), as well as in assessing the effects of
 CC exogenous compound on the expression of nucleic acid and amino acid
 CC sequences of nucleic acid-associated proteins. The invention is useful in
 CC gene therapy. The present sequence is human NAAP-9
 XX
 SQ Sequence 1256 AA;
 Query Match 99.8%; Score 4896; DB 5; length 1256;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 926; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 121 KSEKQLECLTHCIKVPMLVLDPALPANITLKDLPSLYSFHSASDIPNVAKPKNSTNV 180
 Qy 181 SVVVPDSTKQVEDAHSGLLKGNRQYWRGYLTDDXEVPGVLVMQDLAFUSGPPPKET 240
 Db 181 SVVVPDSTKQVEDAHSGLLKGNRQYWRGYLTDDXEVPGVLVMQDLAFUSGPPPKET 240
 Qy 241 NOLKTKLPENTLSKVKLLQLYSEASVALKLNPKPOEANKTKNMITDGKELTISA 300
 Db 241 NOLKTKLPENTLSKVKLLQLYSEASVALKLNPKPOEANKTKNMITDGKELTISA 300
 Qy 301 YLLMDSAISQSKODEISASREDENEELRYSLSIERHAPWVRNIPIVYNGQIPSMNL 360
 Db 301 YLLMDSAISQSKODEISASREDENEELRYSLSIERHAPWVRNIPIVYNGQIPSMNL 360
 Qy 361 DNPRTVTVHODVFRNLSHLPTSSPAISHIRIEGLSQKFTYLANDVWFGKDWPPDP 420
 Db 361 DNPRTVTVHODVFRNLSHLPTSSPAISHIRIEGLSQKFTYLANDVWFGKDWPPDP 420
 Qy 421 YSHSKGQKYLTPVPNCAEGCPSWIKDGYCDKACNNSACDWDGDCSGNSGSGRYIAG 480
 Db 421 YSHSKGQKYLTPVPNCAEGCPSWIKDGYCDKACNNSACDWDGDCSGNSGSGRYIAG 480
 Qy 481 GGGTGSIGVGHFWQFGGINSVYCNQGCANSMYADKFCDOACNVISCGFDADGCGDHF 540
 Db 481 GGGTGSIGVGHFWQFGGINSVYCNQGCANSMYADKFCDOACNVISCGFDADGCGDHF 540
 Qy 541 HELYKVILPNQTHYIIPKGECLPYFSPAFAVARGVAGASDPIIRHASIANKKTIHL 600
 Db 541 HELYKVILPNQTHYIIPKGECLPYFSPAFAVARGVAGASDPIIRHASIANKKTIHL 600
 Qy 601 IMHSGNATTHFNLTQNTNDEEFPMQITVEVTEEGPLNSTAGYENLVSPITLLP 660
 Db 601 IMHSGNATTHFNLTQNTNDEEFPMQITVEVTEEGPLNSTAGYENLVSPITLLP 660
 Qy 661 EARLPEDPKPKRPFKRDVNSRBAOEVRKIPLVNISILPKRAQISLNTLDLQLEH 720
 Db 661 EARLPEDPKPKRPFKRDVNSRBAOEVRKIPLVNISILPKRAQISLNTLDLQLEH 720
 Qy 721 GDITLKGNVLSKSAALRFLMNSQAHAKIRNOAITDETNDLSVAPQEKVHKSILPNSLG 780
 Db 721 GDITLKGNVLSKSAALRFLMNSQAHAKIRNOAITDETNDLSVAPQEKVHKSILPNSLG 780
 Qy 781 VSRRLQRLTPPAVSVKNGHDQGNPDLDTTARFRVETHQKTIGSVYTKKPSLIY 840
 Db 781 VSRRLQRLTPPAVSVKNGHDQGNPDLDTTARFRVETHQKTIGSVYTKKPSLIY 840
 Qy 841 PLSQMTKEKKIKGKEKENSMEENAHNIGTEVLLGRLOHYTDSYIGFLPWEKKKTF 900
 Db 841 PLSQMTKEKKIKGKEKENSMEENAHNIGTEVLLGRLOHYTDSYIGFLPWEKKKTF 900
 Qy 901 QDILDEBSIKQLAVFTDSKNTGRQLK 928
 Db 901 QDILDEBSIKQLAVFTDSKNTGRQLK 928
 RESULT 6
 ADD27810
 ID ADD27810 standard; protein; 1199 AA.
 AC ADD27810;
 XX
 DT 15-JUN-2004 (first entry)
 XX
 DE Soluble human GlcNAc-phosphotransferase.
 XX
 KM human; protein phosphorylation; soluble GlcNAc-phosphotransferase;
 KM UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease;
 KM enzyme.
 OS Synthetic.
 OS Homo sapiens.
 XX
 FN US2003119088-A1.

```

XX 26-JUN-2003.
PD
XX 21-DEC-2001; 2001US-00023888.
PF
XX 21-DEC-2001; 2001US-00023888.
PR
XX (NOVA-) NOVAZYME PHARM INC.
PA
XX Canfield W, Kudo M;
PI
XX WPI; 2003-801323/75.
DR
XX N-PSDB; ADD27809.
XX
XX Phosphorylating a protein for treating a patient suffering from a
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
PT protein.
XX
XX Claim 3; SEQ ID NO 2; 55pp; English.
XX
XX The invention relates to a method of phosphorylating a protein comprising
XX contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
XX acetylglucosamine) and producing a phosphorylated protein. The method is
XX useful for treating a patient suffering from a lysosomal storage disease
XX e.g. Fabry's disease. The present sequence represents the amino acid
XX sequence of soluble human GlcNAc-phosphotransferase.
XX
XX Sequence 1199 AA:
XX
Query Match 94.4%; Score 4630; DB 7; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 45 SSDQYVLPDSRDNDIAGKSPONRLCLPMPIDVYTYWNGTDLLELKELOVREOMESEQ 104
DB 35 SSDQHVLPDSRDNDIAGKSPONRLCLPMPIDVYTYWNGTDLLELKELOVREOMESEQ 94
QY 105 KAMREILGKNTTEPTKSEKOLECLTHCIVKPMVLDPALPANITLKDVPISLYSFHSA 164
DB 95 KAMREILGKNTTEPTKSEKOLECLTHCIVKPMVLDPALPANITLKDVPISLYSFHSA 154
QY 165 SDIFNVAKKPNSTNVSVVFPDSTKDVEDAHSGLLKNGSRQVWVGYYLTTPKEVPGVLYM 224
DB 155 SDIFNVAKKPNSTNVSVVFPDSTKDVEDAHSGLLKNGSRQVWVGYYLTTPKEVPGVLYM 214
QY 225 QDLAPLGGPPTPEKTNQLKTKLLENLSKXVLLQLYSBASVALKLNPPXDFELNKQT 284
DB 215 QDLAPLGGPPTPEKTNQLKTKLLENLSKXVLLQLYSBASVALKLNPPXDFELNKQT 274
QY 285 KKNMTIDGKELITSPAYLLMDLSAISQSKODEDISASREDNEELRYSLSRIERAPVVR 344
DB 275 KKNMTIDGKELITSPAYLLMDLSAISQSKODEDISASREDNEELRYSLSRIERAPVVR 334
QY 345 NIFIVTNGQIPSMNLNDNPRVTIVTHQVFRNLISHLPTFSSPAIESHTRIEGLSOKFIY 404
DB 335 NIFIVTNGQIPSMNLNDNPRVTIVTHQVFRNLISHLPTFSSPAIESHTRIEGLSOKFIY 394
QY 405 LNDVWVFGKQVWPDFFYSKSGQKYLLTWVPVNCABCGPBGSKYIKGICDKAKNNSACMD 464
DB 395 LNDVWVFGKQVWPDFFYSKSGQKYLLTWVPVNCABCGPBGSKYIKGICDKAKNNSACMD 454
QY 465 GDDCGSGSGSRYIAGGGGTGSIQVGHPMQFGGINSYSCNGGACNMLDKDFDQACN 524
DB 455 GDDCGSGSGSRYIAGGGGTGSIQVGHPMQFGGINSYSCNGGACNMLDKDFDQACN 514
QY 525 VLSGCFDAGDCGDHFEHLYKVIILPNQTHYIIPYGECLPYFSFAEVAKRGVEGAYSDNP 584
DB 515 VLSGCFDAGDCGDHFEHLYKVIILPNQTHYIIPYGECLPYFSFAEVAKRGVEGAYSDNP 574
QY 585 IIRHASIAKMKTIHLIHMSSGNAATTIHFNLFQONTNDEEFMOJTTVVDREBKJNST 644
DB 575 IIRHASIAKMKTIHLIHMSSGNAATTIHFNLFQONTNDEEFMOJTTVVDREBKJNST 634

```

```

QY 645 AOKGYENIVSPITLLPEAILFEDIPKPKRPKXBDVNSRRAOEYKIVLNIISLP 704
DB 635 AOKGYENIVSPITLLPEAILFEDIPKPKRPKXBDVNSRRAOEYKIVLNIISLP 694
QY 705 KDAQLSINTLDQLEHGDITLKGYNLSKALLRSPFMSQAKIKNOAIITDETNDSLVA 764
DB 695 KDAQLSINTLDQLEHGDITLKGYNLSKALLRSPFMSQAKIKNOAIITDETNDSLVA 754
QY 765 PQKQVYHKSILPNSLGVSERLQRLTPPANSVKNHDOQGNPPLLETARFRVETHOK 824
DB 755 PQKQVYHKSILPNSLGVSERLQRLTPPANSVKNHDOQGNPPLLETARFRVETHOK 814
QY 825 TIGGANTKEKPESLIVPLESOMTKKIKYKKEKENSMEENANENHIGTEVLGKRLQHY 884
DB 815 TIGGANTKEKPESLIVPLESOMTKKIKYKKEKENSMEENANENHIGTEVLGKRLQHY 874
QY 885 TDSYLGLPFWKPKKYQDILLDEESLKTOLAYFTTSKNTGR 925
DB 875 TDSYLGLPFWKPKKYQDILLDEESLKTOLAYFTTSKNTGR 915

RESULT 7
ABM01487
ID ABM01487 standard; protein; 1199 AA.
XX
AC ABM01487;
XX
DE 15-JAN-2004 (first entry)
XX
DE N-acetylglucosamine-1(GlcNAc)-phosphotransferase.
XX
KM Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
KM lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
KM gastrointestinal enzyme; lectin resistant cell; deoxymannojirimycin;
KM kifunensine; glycosylation inhibition.
XX
OS Unidentified.
XX
PN US2003124652-A1.
XX
PD 03-JUL-2003.
XX
PE 21-DEC-2001; 2001US-00023889.
XX
PR 21-DEC-2001; 2001US-00023889.
XX
PA (NOVA-) NOVAZYME PHARM INC.
XX
PI Canfield WM;
XX
DR WPI; 2003-810984/76.
DR N-PSDB; AAD62490.
XX
PT Producing a high mannose glycoprotein for treating lysosomal storage
PT disease, comprises culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
XX
PS Claim 8; Page 10-13; 46pp; English.
XX
XX The invention relates to a method for producing a high mannose
XX glycoprotein. The method comprises: introducing and expressing a
XX polynucleotide encoding a glycoprotein into a mammalian cell; culturing
XX the cell in the presence of a lectin to obtain a lectin resistant cell;
XX isolating the cell; culturing the cell in the presence of
XX deoxymannojirimycin and kifunensine to inhibit glycosylation of the
XX glycoprotein; and collecting the glycoprotein. The invention is useful in
XX gene therapy. The method is useful for producing a high mannose
XX glycoprotein in a complex carbohydrate deficient cell for treating
XX lysosomal storage disease. The present sequence is N-acetylglucosamine-1
XX (GlcNAc)-phosphotransferase
XX
XX Sequence 1199 AA:

```

Query Match 94.4%; Score 4630; DB 7; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SSDQYHVLFDYSYRDNIAGKSFQNRCLPMPIDVYVTVWNGTDLELKELOQVREQMEEQ 104
DB 35 SSDQYHVLFDYSYRDNIAGKSFQNRCLPMPIDVYVTVWNGTDLELKELOQVREQMEEQ 94
QY 105 KAMREILGKNTTEPTTKSEKQLECLTHCIKVPMLVLDPALPANTITLKDVPSLPSFSA 164
DB 95 KAMREILGKNTTEPTTKSEKQLECLTHCIKVPMLVLDPALPANTITLKDVPSLPSFSA 154
QY 165 SDIFNVAKPKNPSTNVSVVFPDSTKDVEDAHSGLLKGNRSQVWVRGYLTTDKVPGVLVM 224
DB 155 SDIFNVAKPKNPSTNVSVVFPDSTKDVEDAHSGLLKGNRSQVWVRGYLTTDKVPGVLVM 214
QY 225 QDLAFISGFPPTFKETNQIKTKLPENLSKVKLLQLYSEASVALIKLNNPKDFQELNKOT 284
DB 215 QDLAFISGFPPTFKETNQIKTKLPENLSKVKLLQLYSEASVALIKLNNPKDFQELNKOT 274
QY 285 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASREFDNEELRYSLSIRHAPWVR 344
DB 275 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASREFDNEELRYSLSIRHAPWVR 334
QY 345 NIFIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIREGLSQKFTY 404
DB 335 NIFIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIREGLSQKFTY 394
QY 405 LNDVWFGKDVWPDDFYSHSKQKYLLTPVPNCABGCGPGSMTKGYCDKACNNSACMD 464
DB 395 LNDVWFGKDVWPDDFYSHSKQKYLLTPVPNCABGCGPGSMTKGYCDKACNNSACMD 454
QY 465 GSDCGSGNSGSRITVIAAGGCTGSIIGVGHFWQFGGINSVSYCNOGCANSMADKFCQACN 524
DB 455 GSDCGSGNSGSRITVIAAGGCTGSIIGVGHFWQFGGINSVSYCNOGCANSMADKFCQACN 514
QY 525 VLSGCFDAGDCGDHFHELYKYLIPNCTHYIIPKGECLPYFSPAIEVAKRGVEGAYSDNP 584
DB 515 VLSGCFDAGDCGDHFHELYKYLIPNCTHYIIPKGECLPYFSPAIEVAKRGVEGAYSDNP 574
QY 585 IIRHASIANKMTIHLIMHSGNNAATTIHNLFTQNTDNEEFMOQITVEYDTRREGPKLNT 644
DB 575 IIRHASIANKMTIHLIMHSGNNAATTIHNLFTQNTDNEEFMOQITVEYDTRREGPKLNT 634
QY 645 AOKGYENIVSPITLLPEAEILFEDIPEKGRPFKPRHDVNSTRAQEBVKIPLVNISLLP 704
DB 635 AOKGYENIVSPITLLPEAEILFEDIPEKGRPFKPRHDVNSTRAQEBVKIPLVNISLLP 694
QY 705 KNAQSLNLTDLQLEHGDITLKGYNLKSALLRSLVMSQAKIKNOAIIIDETNDSIYA 764
DB 695 KNAQSLNLTDLQLEHGDITLKGYNLKSALLRSLVMSQAKIKNOAIIIDETNDSIYA 754
QY 765 POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNHGQONPDLLETTAFRVSTHTQX 824
DB 755 POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNHGQONPDLLETTAFRVSTHTQX 814
QY 825 TIGGNTTKKPPSLIYPLESOMTKKKITGKEKENSMEENAEHNIIVTEVILGKLOHY 884
DB 815 TIGGNTTKKPPSLIYPLESOMTKKKITGKEKENSMEENAEHNIIVTEVILGKLOHY 874
QY 885 TDSYLGFLPMEKKKYFQDLDDEESLKTQLAFTDSKNTGR 925
DB 875 TDSYLGFLPMEKKKYFQDLDDEESLKTQLAFTDSKNTGR 915

RESULT 8
ABW01536
ID ABW01536 standard; protein; 1199 AA.
XX
AC ABW01536;
XX
DT 15-JAN-2004 (first entry)

XX N-acetylglucosamine-1(GlcNAc)-phosphotransferase protein.
DE Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
XX N-acetylglucosamine-1-phosphotransferase; gene therapy; enzyme.
KM
XX
OS
XX
XX US2003124653-A1.
XX
XX 03-JUL-2003.
PD
XX
XX 21-DEC-2001; 2001US-00023890.
PF
XX
XX 21-DEC-2001; 2001US-00023890.
PR
XX
XX (NOVA-) NOVARTIS PHARM INC.
PA
XX
XX Canfield WM;
PI
XX
XX WPI; 2003-810985/76.
DR
XX
XX N-PSDB; AAD62649.
DR
XX
XX Producing a glycoprotein with reduced complex carbohydrates by culturing
FT the lectin resistant mammalian cell expressing the glycoprotein for
PT treating lysosomal storage disease.
PS
XX
XX Claim 8; Page 10-13; 46pp; English.
PS
XX
XX The present invention provides a method of producing a glycoprotein
CC having reduced complex carbohydrates by culturing the lectin resistant
CC mammalian cell expressing the glycoprotein. The method is useful for
CC producing a glycoprotein with reduced complex carbohydrates for treating
CC lysosomal storage disease. The present invention is also useful in gene
CC therapy. The present sequence is N-acetylglucosamine (GlcNAc)-
CC phosphotransferase protein
XX
SQ Sequence 1199 AA;

Query Match 94.4%; Score 4630; DB 7; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SROCYHVLFDYSYRDNIAGKSFQNRCLPMPIDVYVTVWNGTDLELKELOQVREQMEEQ 104
DB 35 SROCYHVLFDYSYRDNIAGKSFQNRCLPMPIDVYVTVWNGTDLELKELOQVREQMEEQ 94
QY 105 KAMREILGKNTTEPTTKSEKQLECLTHCIKVPMLVLDPALPANTITLKDVPSLPSFSA 164
DB 95 KAMREILGKNTTEPTTKSEKQLECLTHCIKVPMLVLDPALPANTITLKDVPSLPSFSA 154
QY 165 SDIFNVAKPKNPSTNVSVVFPDSTKDVEDAHSGLLKGNRSQVWVRGYLTTDKVPGVLVM 224
DB 155 SDIFNVAKPKNPSTNVSVVFPDSTKDVEDAHSGLLKGNRSQVWVRGYLTTDKVPGVLVM 214
QY 225 QDLAFISGFPPTFKETNQIKTKLPENLSKVKLLQLYSEASVALIKLNNPKDFQELNKOT 284
DB 215 QDLAFISGFPPTFKETNQIKTKLPENLSKVKLLQLYSEASVALIKLNNPKDFQELNKOT 274
QY 285 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASREFDNEELRYSLSIRHAPWVR 344
DB 275 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASREFDNEELRYSLSIRHAPWVR 334
QY 345 NIFIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIREGLSQKFTY 404
DB 335 NIFIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIREGLSQKFTY 394
QY 405 LNDVWFGKDVWPDDFYSHSKQKYLLTPVPNCABGCGPGSMTKGYCDKACNNSACMD 464
DB 395 LNDVWFGKDVWPDDFYSHSKQKYLLTPVPNCABGCGPGSMTKGYCDKACNNSACMD 454
QY 465 GSDCGSGNSGSRITVIAAGGCTGSIIGVGHFWQFGGINSVSYCNOGCANSMADKFCQACN 524

QY 481 GGGTSGIGVGHWMQFGGINSVSYCNQGCANSWLADKFCDOACNTLSCGFDAGCGQDHF 540
DB 435 -----FGGGINSVSYCNQGCANSWLADKFCDOACNTLSCGFDAGCGQDHF 480
QY 541 HELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASTANKKTIHL 600
DB 481 HELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASTANKKTIHL 540
QY 601 IMHSGNATTHIHNLTFOQNTNDEEFMOQITVEVDTRBGPKNSTAKGYENLVSPITLLP 660
DB 541 IMHSGNATTHIHNLTFOQNTNDEEFMOQITVEVDTRBGPKNSTAKGYENLVSPITLLP 600
QY 661 EAEILFEDIPKEKRPFKRHDVNSTRAQOEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
DB 601 EAEILFEDIPKEKRPFKRHDVNSTRAQOEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 660
QY 721 GDITLKGYNSKALLRSFLMNSQAKIKNOAIITDETNDLSVAPQEKVAKSILPNSLG 780
DB 661 GDITLKGYNSKALLRSFLMNSQAKIKNOAIITDETNDLSVAPQEKVAKSILPNSLG 720
QY 781 VSERLQRLTFPAVSVKVNGHDQGNPDLLETTAFRVEVTHQKTIGAVTKEKPPSLIV 840
DB 721 VSERLQRLTFPAVSVKVNGHDQGNPDLLETTAFRVEVTHQKTIGAVTKEKPPSLIV 780
QY 841 PLESQWTEKKITTGKEKENSMEENAHNIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900
DB 781 PLESQWTEKKITTGKEKENSMEENAHNIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 840
QY 901 QDLDBESLKTQLAYFTDSCKTGRQK 928
DB 841 QDLDBESLKTQLAYFTDSCKTGRQK 900
RESULT 10
ADA54824
ID ADA54824 standard, protein, 846 AA.
XX
AC ADA54824;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 2392.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nocitropi;
KW Gene therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
XX
OS Homo sapiens.
XX
FN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
XX
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S,
PI Yamamoto U, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
PI Seki N, Yoshikawa T, Otsuka W, Nagahari K, Masuho Y;
XX
DR WPI; 2003-395539/38.
XX
DR N-PSDB; ADA53185.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 14; SEQ ID NO 2392; 205bp; English.

XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 846 AA;
Query Match 91.1%; Score 4468; DB 6; Length 846;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 844; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLFKLQRCQYTCLSHRGGLVYVCFGLGVVYTVISAFQFGVYLEMSDQYHVLFDSPRDN 60
DB 1 MLFKLQRCQYTCLSHRGGLVYVCFGLGVVYTVISAFQFGVYLEMSDQYHVLFDSPRDN 60
QY 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELKEIQVREQMEBEOKAMRZILGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELKEIQVREQMEBEOKAMRZILGKNTTEPTK 120
QY 121 KSEKQECCLTTCICIKPMLVLPALPANITLKQVSELVYFSASDIENVAKPKPSTNV 180
DB 121 KSEKQECCLTTCICIKPMLVLPALPANITLKQVSELVYFSASDIENVAKPKPSTNV 180
QY 181 SVVVPDSKQVEDAHSGLLKGNRQTVWRGYLTTDKEVPGLVIMODLAFSGPPTFEKT 240
DB 181 SVVVPDSKQVEDAHSGLLKGNRQTVWRGYLTTDKEVPGLVIMODLAFSGPPTFEKT 240
QY 241 NQKTKLPENLSKRYKLLQYSPASVALLKLNPKDFQELNKKYKKNMTIDGKELTISPA 300
DB 241 NQKTKLPENLSKRYKLLQYSPASVALLKLNPKDFQELNKKYKKNMTIDGKELTISPA 300
QY 301 YLLMDLSAISQKODEISASRPEDNBEIYSRSTIERHAPWRNFIYVNGQIPSMNL 360
DB 301 YLLMDLSAISQKODEISASRPEDNBEIYSRSTIERHAPWRNFIYVNGQIPSMNL 360
QY 361 DNPRVTIVTHQDVFRLSHLPTFSSPAISHIRIGLSQKFIYLLNDVWFGDVPDDF 420
DB 361 DNPRVTIVTHQDVFRLSHLPTFSSPAISHIRIGLSQKFIYLLNDVWFGDVPDDF 420
QY 421 YSHSKGQKYLITWVPVNCABGCGSWIKQYCPKACNNSACMDWDGDCGNSGGSRYIAG 480
DB 421 YSHSKGQKYLITWVPVNCABGCGSWIKQYCPKACNNSACMDWDGDCGNSGGSRYIAG 480
QY 481 GGGTSGIGVGHWMQFGGINSVSYCNQGCANSWLADKFCDOACNTLSCGFDAGCGQDHF 540
DB 481 GGGTSGIGVGHWMQFGGINSVSYCNQGCANSWLADKFCDOACNTLSCGFDAGCGQDHF 540
QY 541 HELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASTANKKTIHL 600
DB 541 HELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASTANKKTIHL 600
QY 601 IMHSGNATTHIHNLTFOQNTNDEEFMOQITVEVDTRBGPKNSTAKGYENLVSPITLLP 660
DB 601 IMHSGNATTHIHNLTFOQNTNDEEFMOQITVEVDTRBGPKNSTAKGYENLVSPITLLP 660
QY 661 EAEILFEDIPKEKRPFKRHDVNSTRAQOEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
DB 661 EAEILFEDIPKEKRPFKRHDVNSTRAQOEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
QY 721 GDITLKGYNSKALLRSFLMNSQAKIKNOAIITDETNDLSVAPQEKVAKSILPNSLG 780
DB 721 GDITLKGYNSKALLRSFLMNSQAKIKNOAIITDETNDLSVAPQEKVAKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKVNGHDQGNPDLLETTAFRVEVTHQKTIGAVTKEKPPSLIV 840
DB 781 VSERLQRLTFPAVSVKVNGHDQGNPDLLETTAFRVEVTHQKTIGAVTKEKPPSLIV 840
QY 841 PLESQW 846
DB 841 PLESQW 846

RESULT 11
 AB07381
 ID AB07381 standard; protein; 1459 AA.
 XX
 AC AB07381;
 XX
 DT 28-JAN-2003 (first entry)
 XX
 DE Human protein NOV12.
 XX
 KW Human; NOVX; cardiomyopathy; atherosclerosis; cancer; hypertension;
 KW diabetes; inflammation; autoimmune disorder; allergy; blood disorder;
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
 KW immunoglobulin A nephropathy; cirrhosis; arthritis; Alzheimer's disease;
 KW Parkinson's disease; goitre; infection; stroke; muscular dystrophy;
 KW epilepsy; wasting disorder; neurogenesis; cell differentiation;
 KW cell proliferation; haematopoiesis; wound healing; angiogenesis;
 KW gene therapy; single nucleotide polymorphism; SNP.
 XX
 OS Homo sapiens.
 XX
 PN WO200285922-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 11-APR-2002; 2002WO-US011634.
 XX
 PR 23-APR-2001; 2001US-0285748P.
 PR 24-APR-2001; 2001US-0286068P.
 PR 25-APR-2001; 2001US-0286292P.
 PR 03-MAY-2001; 2001US-0286334P.
 PR 16-MAY-2001; 2001US-0291241P.
 PR 14-SEP-2001; 2001US-0322284P.
 XX
 PA (CURA-) CURAGEN CORP.
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Pena CE, Guo X, Shinkets RA, Padigaru M, Kekuda R, Spytek KA,
 PI Mehreban F, Topper JN, Malyankar UM, Wasserman S, Edinger S,
 PI Smithson G, Gunther E, Komuves L;
 XX
 DR WPI; 2003-058712/05.
 XX
 DR N-PSDB; ABX10235.
 PT
 PT New NOVX polypeptides and nucleic acids, useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis or diabetes, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.
 PT
 XX
 PS Claim 1; Page 94-95; 301pp; English.
 CC
 CC The invention relates to an isolated polypeptide comprising any one of 17
 CC human NOVX (1-9, 10a, 10b, 11-16) appearing as AB07389-AB07385, a
 CC mature form of 1c, or a variant of them, where one or more residues of
 CC the variant differs in not more than 15 % from the residues of the
 CC sequence of them and their encoding polynucleotides appearing as ABX10223
 CC -ABX10239. Also included are NOVX expression vectors, transformed cells,
 CC antibodies, identifying an agent that binds to or modulates the
 CC expression or activity of NOVX and screening for a modulator of activity
 CC or of latency or predisposition to a NOVX-associated disorder. The NOVX
 CC polypeptides, polynucleotides and antibodies are useful in manufacturing
 CC a medicament for treating or preventing a syndrome associated with NOVX-
 CC associated disorder, such as cardiomyopathy, atherosclerosis, cancer,
 CC hypertension, diabetes, inflammation, autoimmune disorders, allergies,
 CC blood disorders, acquired immunodeficiency syndrome (AIDS), obesity,
 CC asthma, immunoglobulin (Ig) A nephropathy, cirrhosis, arthritis,
 CC Alzheimer's disease, Parkinson's disease, goitre, infections (e.g.
 CC bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy, and
 CC other wasting disorders associated with chronic diseases. The nucleic
 CC acids and polypeptides may also be used as targets for the identification
 CC of small molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and

angiogenesis, in gene therapy, in generation of antibodies that bind
 CC immunospecifically to NOVX substances for use in therapeutic or
 CC diagnostic methods. The nucleic acids are further used as hybridisation
 CC probes, in chromosome mapping, tissue typing, preventive medicine, and
 CC pharmacogenomics. The polypeptides are also useful as vaccines. The
 CC present sequence represents a NOVX protein
 XX
 XX Sequence 1459 AA;
 SO
 Query Match 89.5%; Score 4391; DB 6; Length 1459;
 Best Local Similarity 73.0%; Pred. No. 0;
 Matches 884; Conservative 13; Mismatches 28; Indels 286; Gaps 13;
 QY 1 MLEKLLQRTYTCLSRHYGLVYCFGLGVNTIVSAFQGE----- 39
 Db 1 MLEKLLQRTYTCLSRHYGLVYCFGLGVNTIVSAFQGEVARDPAKAPIVHRTAPTK 60
 QY 40 ----- 39
 Db 61 NHPAQNVDAAVEKSGIRRGKNGRAVSLQDMPGTRGCANFTPAFCHDCKEVSQKRL 120
 QY 40 -----VLEMSRDQYVHFLDSYRD 58
 Db 121 YILQNGHMLTDMGWTWLLHSGSLIQGPASBPGCVLLKAKVVLWEMSRDQYVHFLDSYRD 180
 QY 59 NINAGSFQNRRLCPMPIDVYVTVWNGTDELKELQVREQMEBEOKANRELIGKTTTP 118
 Db 181 NINAGSFQNRRLCPMPIDVYVTVWNGTDELKELQVREQMEBEOKARELIGKTTTP 240
 QY 119 TKSEKQELCLTHCICKVEMLVDPALPANITLKDVPSPYSPFHSASDIFNVAKRNPT 178
 Db 241 TKSEKQELCLTHCICKVEMLVDPALPANITLKDVPSPYSPFHSASDIFNVAKRNPT 300
 QY 179 NVSVVVPDSTKD-----VEDHSGLLKGN-----ROTWT----- 208
 Db 301 NVSVVVPDSTKDGLTQKVTPEWKCEGEVSNANIMWKTDLGSPRRPLPMPVALLPBR 360
 QY 209 -----RGYLTDEKVEPGLV- MODLAFIAGSP 234
 Db 361 AQLSSALQILTRPVSDPANTSYLEKIDTPILRGTA--KEVPBETGLQSLSAHCLP 417
 QY 235 -----PTF--KET-----NQLKT--KL 247
 Db 418 TDLIRKNELRDEVKMYKODPSILHTKTCFLREQVESMGESYKSEENIKELTKGSKV 477
 QY 248 PEN-----LSKKYKQLQLYS-----EASVALLKLNPKDFOELNKQTK 285
 Db 478 ENISTDELSSESDLEIDNEAVIEPDIDSPEQMGDEASVALKLNPKDFOELNKQTK 537
 QY 286 KMMTIDGKELTISPAVILMDLSAISQSKODEDISASRPDNEELRYSLSIERHAPVYN 345
 Db 538 KMMTIDGKELTISPAVILMDLSAISQSKODEDISASRPDNEELRYSLSIERHAPVYN 597
 QY 346 IFVTVNGQIPSWLMDNPRVTIYTHQDVPRANISHLPFTSSPAIBSHIRIEGLSKFTYL 405
 Db 598 IFVTVNGQIPSWLMDNPRVTIYTHQDVPRANISHLPFTSSPAIBSHIRIEGLSKFTYL 657
 QY 406 NDVWVGKQVWPPDFFSHSKGQKVVLTWVPVCAEGCPGSMWTKDGYCDKACNNSACMDMG 465
 Db 658 NDVWVGKQVWPPDFFSHSKGQKVVLTWVPVCAEGCPGSMWTKDGYCDKACNNSACMDMG 717
 QY 466 GDCSGNSGSGSRITAGGGGTGSGVGHPMWFGGGINSVSYCNQGCANSLADYFCQACNV 525
 Db 718 GDCSGNSGSGSRITAGGGGTGSGVGHPMWFGGGINSVSYCNQGCANSLADYFCQACNV 777
 QY 526 LSCGFPAGDQGO-----DHFEELYYVILLPNQTHYII 557
 Db 778 LSCGFPAGDQGOENSGSKRTEKCEPKYKKIMLFPFLDHFHLLKYVILLPNQTHYII 837
 QY 558 PKGECIPYSPFAVAVARVEGAYSDNPITIRHASINKKWTIHLIMSGMNAATTIHPNLT 617
 Db 838 PKGECIPYSPFAVAVARVEGAYSDNPITIRHASINKKWTIHLIMSGMNAATTIHPNLT 897

QY 618 QNTNDEEFKMOITVEVDTRBEPKLNSTAKGYENLVSPITLLPEAILFEDIPKREPKK 677
 DB 898 QNTNDEEFKMOITVEVDTRBEPKLNSTAKGYENLVSPITLLPEAILFEDIPKREPK 957
 QY 678 FRHDVNSTRAQOEVEKIPLVNISLLPKDAQSLNTLDLOEHGDIITLKGYNLSKALLR 737
 DB 956 FRHDVNSTRAQOEVEKIPLVNISLLPKDAQSLNTLDLOEHGDIITLKGYNLSKALLR 1017
 QY 738 STLMNSQAKIKNOAIIIDETNDSLVAPQEKVHNSIIPNSIGVSEIRQRLTFPAVSXKV 797
 DB 1018 STLMNSQAKIKNOAIIIDETNDSLVAPQEKVHNSIIPNSIGVSEIRQRLTFPAVSXKV 1077
 QY 798 NSHDGQNPPLDLETTARFRVETHTQKTIGVNTKEKPPSLIVPLESQWTEKKTGKEX 857
 DB 1078 NSHDGQNPPLDLETTARFRVETHTQKTIGVNTKEKPPSLIVPLESQWTEKKTGKEX 1137
 QY 858 ENSRMEENAEHIGVTEVLGRKLOHYTDSYIGFLPWKKKYFQDLDEESLKTOLAYF 917
 DB 1138 ENSRMEENAEHIGVTEVLGRKLOHYTDSYIGFLPWKKKYFQDLDEESLKTOLAYF 1197
 QY 918 TDSKNTGRQLK 928
 DB 1198 TDSKNTGRQLK 1208

RESULT 12
 ABR61384
 ID ABR61384 standard; protein; 908 AA.
 XX ABR61384;
 DT 01-AUG-2003 (first entry)
 XX
 DE Murine GlcNAc-phosphotransferase alpha-subunit.
 XX
 KW Mouse; N-acetylglucosamine-1-phosphotransferase; nephrotropic;
 KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;
 KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;
 KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;
 KW lysosomal storage disease; enzyme; alpha-subunit.
 XX
 OS Mus musculus.
 XX
 PN US6537785-B1.
 PD 25-MAR-2003.
 XX
 PF 10-AUG-2000; 2000US-00636077.
 XX
 PR 14-SEP-1999; 99US-0153831P.
 PA (GEN2-) GENZYME GLYCOBIOLOGY RES INST INC.
 XX
 PI Canfield MW;
 XX
 DR WPI; 2001-290356/30.
 DR N-PSDB; ACC81007.
 XX
 PT Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing PT phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.
 PS
 XX Disclosure; Page 44-47; 62pp; English.
 CC The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (II). The protein of the invention has CC nephrotropic activity, and may be useful in enzyme replacement therapy. A CC protein of the invention (I), (II) is useful for preparing a CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase CC comprising a terminal mannose-6-phosphate, is useful for treating a

CC patient suffering from a lysosomal storage disease. The present sequence
 CC is used in the exemplification of the invention
 XX
 SQ Sequence 908 AA;
 Query Match 79.4%; Score 3894; DB 4; Length 908;
 Best Local Similarity 79.9%; Pred. No. 3, 6e-305;
 Matches 742; Conserved 61; Mismatches 104; Indels 22; Gaps 7;
 QY 1 MLFKLQRTYTCLSRHYGLVYCFGLGVVVTIVSAFQFGVYVLEMSRDQHYVLFDSYRDN 60
 DB 1 MLFKLQRTYTCLSRHYGLVYCFGVVVTIVSAFQFGVYVLEMSRDQHYVLFDSYRDN 60
 QY 61 AGSFQNRCLPMPIDVVYTWVNGTDLLELLEKLOQVREBEMERKMRZLIGNTEPRK 120
 DB 61 AGSFQNRCLPMPIDVVYTWVNGTDLLELLEKLOQVREBEMERKMRZLIGNTEPRK 120
 QY 121 KSEKQLECLLTHCIVKPMVLDPALPANTLTKDVPSLYPSFASDIFVNAKXPNSTNV 180
 DB 121 KSEKQLECLLTHCIVKPMVLDPALPATITLKDPLTLYSPFASDMFVNAKXPNSTNV 180
 QY 181 SVVVPDSTKQVEDNAHSGILKGNRQTVWRGYLTIDKEVGVLYMQDLATLSGPPPTFKET 240
 DB 181 SVVVPDSTKQVEDNAHSGILKGNRQTVWRGYLTIDKEVGVLYMQDLATLSGPPPTFKET 240
 QY 181 PVVVPDSTKQVEDNAHSGILKGNRQTVWRGYLTIDKEVGVLYMQDLATLSGPPPTFKET 240
 DB 181 PVVVPDSTKQVEDNAHSGILKGNRQTVWRGYLTIDKEVGVLYMQDLATLSGPPPTFKET 240
 QY 241 NQKTKLPEN-LSSKVKLLQVSEASVALKLKNNPKDFOLNKQTKKNTIDKELTIS 299
 DB 241 NQKTKLPEN-LSSKVKLLQVSEASVALKLKNNPKDFOLNKQTKKNTIDKELTIS 299
 QY 241 SGLKTKLPKAPFLKIKLRLYSEASVALKLKNNPKDFOLNKQTKKNTIDKELTIS 300
 DB 241 SGLKTKLPKAPFLKIKLRLYSEASVALKLKNNPKDFOLNKQTKKNTIDKELTIS 300
 QY 300 AYLLMDLSAISQKODEDISASFEDNEELRYSLSIERHAPVWRNIFVTNGQIPSWLN 359
 DB 301 AYLLMDLSAISQKODEDISASFEDNEELRYSLSIERHAPVWRNIFVTNGQIPSWLN 360
 QY 360 LDNPRVYIVTHQVFNLSHLPTFESSPAIESHHRLEGISQKTYLNDQVPEKDVPPD 419
 DB 361 LDNPRVYIVTHQVFNLSHLPTFESSPAIESHHRLEGISQKTYLNDQVPEKDVPPD 420
 QY 420 FYSHSKQKXYLLTPVPCACGGPGSMIDKQYCDXACNNASACDWDGDCSGNSGSGRYTA 479
 DB 421 FYSHSKQKXYLLTPVPCACGGPGSMIDKQYCDXACNNASACDWDGDCSGNSGSGRYTA 480
 QY 480 GGGGTGSIYGVHMGWCGGGINSYCNOCANASWLADKFCDOACNVLSCGPDGDCGQH 539
 DB 481 GGGGTGSIYGVHMGWCGGGINSYCNOCANASWLADKFCDOACNVLSCGPDGDCGQH 540
 QY 540 FHELYKVIILPNTQHYIIPKGECLPYFSFAVAGVEGAYSDNPIIRHASTANKKTH 599
 DB 541 FHELYKVIILPNTQHYIIPKGECLPYFSFAVAGVEGAYSDNPIIRHASTANKKTH 600
 QY 600 LIMSQGNATTFINLTFOYNTNDEFRMQITVAVDTRBEPKLNSTAKGYENLVSPITLL 659
 DB 601 LIMSQGNATTFINLTFOYNTNDEFRMQITVAVDTRBEPKLNSTAKGYENLVSPITLL 660
 QY 660 PEAILFEDIPKREPKFRHDVNSTRAQOEVEKIPLVNISLLPKDAQSLNTLDLOE 719
 DB 661 PEAILFEDIPKREPKFRHDVNSTRAQOEVEKIPLVNISLLPKDAQSLNTLDLOE 720
 QY 720 HGDITLKGYNLSKALLRSFLMNSQAKIKNOAIIIDETNDSLVAPQEKVHNSIIPNSL 779
 DB 721 HGDITLKGYNLSKALLRSFLMNSQAKIKNOAIIIDETNDSLVAPQEKVHNSIIPNSL 780
 QY 780 GVESEIRQRLTFPAVSXKVNCHDQGNPDLLETTARFRVETHTQKTIGVNTKEKPPSL 839
 DB 778 GVESEIRQRLTFPAVSXKVNCHDQGNPDLLETTARFRVETHTQKTIGVNTKEKPPSL 832
 QY 840 VPLESQWTEKKTGKEXENSRMEENAEHIGVTEVLGRKLOHYTDSYIGFLPWKKKY 899
 DB 841 VPLESQWTEKKTGKEXENSRMEENAEHIGVTEVLGRKLOHYTDSYIGFLPWKKKY 899
 QY 899 VPESHLP-----KEESDRAEGNA--VPKELVPRRLQ--QNVPGFLPWKKKY 879
 DB 899 VPESHLP-----KEESDRAEGNA--VPKELVPRRLQ--QNVPGFLPWKKKY 879

RESULT 13
ADD27817 standard; protein; 908 AA.
XX
AC ADD27817;
XX
DT 15-JAN-2004 (first entry)
XX
DE GlcNAc-phosphotransferase associated protein #1.
XX
KM mouse; protein phosphorylation; soluble GlcNAc-phosphotransferase;
KM UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.
XX
OS Mus musculus.
XX
PN US2003119088-A1.
XX
PD 26-JUN-2003.
XX
PF 21-DEC-2001; 2001US-00023888.
XX
PR 21-DEC-2001; 2001US-00023888.
XX
PA (NOVA-) NOVAZYME PHARM INC.
XX
PI Canfield W, Kudo M;
XX
DR WPI; 2003-801323/75.
XX
PT Phosphorylating a protein for treating a patient suffering from a
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
PT protein.
XX
PS Disclosure; SEQ ID NO 9; 55pp; English.
XX
CC The invention relates to a method of phosphorylating a protein comprising
CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
CC acetylglucosamine) and producing a phosphorylated protein. The method is
CC useful for treating a patient suffering from a lysosomal storage disease
CC e.g. Fabry's disease. The present sequence represents the amino acid
CC sequence of a GlcNAc-phosphotransferase associated protein.
XX
SQ Sequence 908 AA;
Query Match 79.4%; Score 3894; DB 7; Length 508;
Best Local Similarity 79.9%; Pred. No. 3,6e-305;
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;
QY 1 MLEFKLQROTYTCLSHRYGLVYCFVGAVVTVISAFQFGVWLEMSRDQYHLPFSYRNI 60
DB 1 MLEFKLQROTYTCLSHRYGLVYCFVGAVVTVISAFQFGVWLEMSRDQYHLPFSYRNI 60
QY 61 AKGSFQNRCLCPMPIDVYTVVNGTDLLELKELOQVRBQMEBQKAMEIIGKNTTEPTK 120
DB 61 AKGSFQNRCLCPMPIDVYTVVNGTDLLELKELOQVRBQMEBQKAMEIIGKNTTEPTK 120
QY 121 KSEKOLECLTHCIVPMVMDPALPANITLKDVPALSPFSHSDIRNVAKPKNPSTNV 180
DB 121 KSEKOLECLTHCIVPMVMDPALPANITLKDVPALSPFSHSDIRNVAKPKNPSTNV 180
QY 181 SYVVPDSTKDVEDAHSGLLKGNRSQTVRGVLTDDKEVPGVLMDQLAFISGPPPTFKET 240
DB 181 SYVVPDSTKDVEDAHSGLLKGNRSQTVRGVLTDDKEVPGVLMDQLAFISGPPPTFKET 240
QY 241 NOLTKRLPEN-LSSTVKYLQLYSEASVALLKLNKPKDQELNKQTKKMTIDGELTISP 299
DB 241 NOLTKRLPEN-LSSTVKYLQLYSEASVALLKLNKPKDQELNKQTKKMTIDGELTISP 299
QY 300 AVILMDLSAISQKODEDISASRFEDNEBELRYSLSIERHAPWYANIFIVTNGQIPSWLN 359
DB 301 AVILMDLSAISQKODEDISASRFEDNEBELRYSLSIERHAPWYANIFIVTNGQIPSWLN 360

QY 360 LDNPRVITVHODVFNRLSHLPFSSPAIESHHRHREGSOKFIYLNDDVFGKDVWPD 419
DB 361 LDNPRVITVHODVFNRLSHLPFSSPAIESHHRHREGSOKFIYLNDDVFGKDVWPD 420
QY 420 FYSHSKQKRYLLTWVPNCAEGCPGSMIKDGYCDRAKNSACDWGDCGSGSGRYTA 479
DB 421 FYSHSKQKRYLLTWVPNCAEGCPGSMIKDGYCDRAKNSACDWGDCGSGSGRYTA 480
QY 480 GGGGTSGIGVHWMQGGGINSYCNOCANSMWADKFCDDACNVLSGFGPAGCGGDH 539
DB 481 RGGGTGNTGAGHWMQGGGINTSYCNOCANSMWADKFCDDACNVLSGFGPAGCGGDH 540
QY 540 FHELYKVILLPNQTHYIIPKGBCLPYFSFAEVAKGVGEAGVSDNPITRASIANKKTIH 599
DB 541 FHELYKVILLPNQTHYIIPKGBCLPYFSFAEVAKGVGEAGVSDNPITRASIANKKTIH 600
QY 600 LNHSGMNAITTHFNITPONTNDEEPRMOITVEVDTREGPKNSTAOKYENLVSPITLL 659
DB 601 LIMPGMNAITTYFENLTONANDEEPRMOITVEVDTREGPKNSTAOKYENLVSPITLL 660
QY 660 PEAEILFEDIPEKREPFKRDVNSTRAQEEVKIPLVNIISLBPDAQLSINTLDLOLE 719
DB 661 PDAVPEFEDVPEKREPFKRRDVAATGFOEEVKIPRNISLLEKRAQVRLSNDLOLE 720
QY 720 HGDITLKGYNLSKSLRLSFLMNSQAKIKNOAIIITDETNSLVAPQEKVHKSILPNSL 779
DB 721 RGDITLKGYNLSKSLRLSFLMNSQAKIKNOAIIITDETNSLVAPQEKVHKSILPNSL 777
QY 780 GVSERLQRLTPPVSQVKNVGHQOGNPPDLFTTARFRVETHTOKTIGAVTKKPPSLI 839
DB 778 AGHREBRTAPAEITYVAGRDHALNPVLETNARL-----AQPTLGTVTKENLSPLI 832
QY 840 VPLESQTKKEKKTIGKEKENSMEENAEHNIQVTEVLLGRKLOHYTDSYGLFLPWEKKY 899
DB 843 VPESHLP-----KEESDRAEGNA--VPKEVLVPRRLQ--QNPFGFLPWEKKY 879
QY 900 FQDLDEESIXTQLAFTDSKNTGRQK 928
DB 880 FQDLDEESIXTQLAFTDSKNTGRQK 908

RESULT 14
ABM01491
ID ABM01491 standard; protein; 908 AA.
XX
AC ABM01491;
XX
DT 15-JAN-2004 (first entry)
XX
DE Mouse protein #1 used to illustrate the method of the invention.
XX
KM Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
KM lysosomal storage disease; gastrointestinal; mouse;
KM lectin resistant cell; deoxymannojirimycin; kifunensine;
KM glycosylation inhibition.
XX
OS Mus musculus.
XX
PN US2003124652-A1.
XX
PD 03-JUL-2003.
XX
PF 21-DEC-2001; 2001US-00023888.
XX
PR 21-DEC-2001; 2001US-00023888.
XX
PA (NOVA-) NOVAZYME PHARM INC.
XX
PI Canfield W;
XX
DR WPI; 2003-810984/76.
XX

PT Producing a high mannose glycoprotein for treating lysosomal storage
PT disease, comprises culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
PS Disclosure; Page 23-25; 46pp; English.
XX
XX
XX The invention relates to a method for producing a high mannose
CC glycoprotein. The method comprises: introducing and expressing a
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
CC the cell in the presence of a lectin to obtain a lectin resistant cell;
CC isolating the cell; culturing the cell in the presence of
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
CC glycoprotein; and collecting the glycoprotein. The invention is useful in
CC gene therapy. The method is useful for producing a high mannose
CC glycoprotein in a complex carbohydrate deficient cell for treating
CC lysosomal storage disease. The present sequence is mouse protein used to
CC illustrate the method of the invention
CC
XX
SQ Sequence 908 AA;

Query Match 79.4%; Score 3894; DB 7; Length 908;
Best Local Similarity 79.9%; Pred. No. 3.6e-305;

Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLEFKLQRTYTCLSHRVGLYVCFGLGVVVTIVSAFQFGEVYLWMSRDQYHLPDSYRDN 60
DB 1 MLEFKLQRTYTCLSHRVGLYVCFGVVVTIVSAFQFGEVYLWMSRDQYHLPDSYRDN 60
QY 61 AGKSFQNRCLPMPIDVYVYVWNGTDLLELKELOQVREOMEBEQKARELIGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYVYVWNGTDLLELKELOQVREHMBEQKARELIGKNTTEPTK 120
QY 121 KSEKQLECHLTHCIVPMVLVDPALPANITLKDVSLYPSFHSADIFNVAKPKNPSTNV 180
DB 121 KSEKQLECHLTHCIVPMVLVDPALPATITLKDLPTLPSFHSADIMNVAKPKNPSTNV 180
QY 121 KSEKQLECHLTHCIVPMVLVDPALPATITLKDLPTLPSFHSADIMNVAKPKNPSTNV 180
DB 181 SVVVFDSSTQVDVAHSGILKNSRQTVNRGYITTDKEVPGVLMODLAFISGFPPTKET 240
QY 181 PVVVFDTTKDVAHAGPFGQQTVDVRAVLTDDKAPAGVLIOGLAFISGFPPTKET 240
QY 241 NOLKTKLPEN-LSSKVKLLQLYSEASVALKLKNPFDQFOELNKQTKMMTIDGKELTISP 299
DB 241 SOLKTKLPKAPLAIKILRLYSEASVALKLKNPFGQELNKQTKMMTIDGKELTISP 300
QY 300 AYLLMDLSAISQKODEDISASRFEDNEELRYSLRSIRHAPVWNIITVNGQIPSWLN 359
DB 301 AYLLMDLSAISQKODEDISASRFEDNEELRYSLRSIRHAPVWNIITVNGQIPSWLN 360
QY 360 LNPVTVITVTHODVFRNLSHLPTSSPAIESHIEHIEGLSOKETIYLVNDVWFGQVWPD 419
DB 361 LNPVTVITVTHODIFQNSHLPTSSPAIESHIEHIEGLSOKETIYLVNDVWFGQVWPD 420
QY 420 FYSHSGQKQVYLTWVPNCABEGCPGSMWKDGYCDRACNNUSACMDGDCSGNSGGSRYIA 479
DB 421 FYSHSGQKQVYLTWVPNCABEGCPGSMWKDGYCDRACNUSPCDMGDCSGNSGGSRYIA 480
QY 480 GGGGCGSISGVHPMGQGGINSVSYCNCGCANSMLADFCDOACNVLSGCFDAGDCGDDH 539
DB 481 RGGGCGNIGAGQHWQFGGINTISYCNCGCANSMLADFCDOACNVLSGCFDAGDCGDDH 540
QY 540 FHELYVILLPQNTYIIPKGECLPYFSFAVAKGVEGAVSDNIIIRHAIANKMTIH 599
DB 541 FHELYVILLPQNTYIIPKGECLPYFSFAVAKGVEGAVSDNIIIRHAIANKMTIH 600
QY 600 LINESGNAATTIHNTFQNTNDEEPKQIYEVDTREGPKLNSIAQGYENLVSPITLL 659
DB 601 LIMPGMNATITTYFNLTIONANDEBEFKIOIAVEVDTREAPKLNSTTORAKAYESLVSPVPL 660
QY 660 PEALIEFEDIPKBPPEKFKRDVNSVTRAOEAVKIPLVNIISLPKQOLSLNTDLE 719
DB 661 PQADVFEEVPEKPEKFKRHDVNAATGFOEBEVIKPRVNISILPKEAQVRLSNDDLE 720
QY 720 HSDITLKGYLSKALLRSFLMNSQHAKIKNOAITTDETNDSLVAPQEKVHKSLIPNSL 779

DB 721 RGDITLKGYLSKALLRSFLMNSQHAKIKNOAITTDETNDSLVAPQEKVHKSLIPNSL 779
QY 780 GVSERLQILTFPAVSVKVNQGHQGNPPLDLETTAFRVEITHQKTIGANVTKEKPSLI 839
DB 778 AGEHSEERWTAPAEITVTKGRDHALNPPEVLETNARL-----AQPTLGVTSKENISPLI 832
QY 840 VPLESQMTKEKKTIGKENSRYEENAEHIGVTEVILGRKLOHYTDSYLGFLPMEKKY 899
DB 833 VPESHLP-----KEESDRAEGNA---VPKSLVPRRLD---QNTPGFLPMEKKY 879
QY 900 FQDLDEESLTKQLAAYFTDSKNTGSLK 928
DB 880 FQDLDEESLTKQLAAYFTDRKHTGSLK 908

RESULT 15

ABW01540
ID ABW01540 standard; protein; 908 AA.
AC
XX ABW01540;
XX
DT 15-JAN-2004 (first entry)

DE Mouse protein #1 used to illustrate the method of the invention.
XX
XX Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
KM N-acetylglucosamine-1-phosphotransferase; gene therapy; mouse.

OS Mus musculus.
XX
XX US2003124653-A1.
XX
XX 03-JUL-2003.
XX
XX 21-DEC-2001; 2001US-00023890.
XX
XX 21-DEC-2001; 2001US-00023890.
XX
XX (NOVA-) NOVAZYME PHARM INC.

PI Canfield WM;
XX
XX WPI; 2003-810965/76.

XX
XX
XX Producing a glycoprotein with reduced complex carbohydrates by culturing
PT the lectin resistant mammalian cell expressing the glycoprotein for
PT treating lysosomal storage disease.

XX Disclosure; Page 23-25; 46pp; English.

XX
XX
XX The present invention provides a method of producing a glycoprotein
CC having reduced complex carbohydrates by culturing the lectin resistant
CC mammalian cell expressing the glycoprotein. The method is useful for
CC producing a glycoprotein with reduced complex carbohydrates for treating
CC lysosomal storage disease. The present invention is also useful in gene
CC therapy. The present sequence is mouse protein used to illustrate the
CC method of the invention
CC
XX
SQ Sequence 908 AA;

Query Match 79.4%; Score 3894; DB 7; Length 908;
Best Local Similarity 79.9%; Pred. No. 3.6e-305;

Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLEFKLQRTYTCLSHRVGLYVCFGLGVVVTIVSAFQFGEVYLWMSRDQYHLPDSYRDN 60
DB 1 MLEFKLQRTYTCLSHRVGLYVCFGVVVTIVSAFQFGEVYLWMSRDQYHLPDSYRDN 60
QY 61 AGKSFQNRCLPMPIDVYVYVWNGTDLLELKELOQVREOMEBEQKARELIGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYVYVWNGTDLLELKELOQVREHMBEQKARELIGKNTTEPTK 120

QY 121 KSEKQECLELTHCIRKVMVLDPALPANITLKDVPSELVPSFHSASDIFPVAKPKPNSTNV 180
 Db 121 KSEKQECLELTHCIRKVMVLDPALPANITLKDVPSELVPSFHSASDIFPVAKPKPNSTNV 180
 QY 181 SVVVPSTYKQVEDAHSGGLKNSRQTVWBGVLTDDKEVGVLMQDLAFLSGFPPTFKET 240
 Db 181 SVVVPSTYKQVEDAHSGGLKNSRQTVWBGVLTDDKEVGVLMQDLAFLSGFPPTFKET 240
 QY 241 NOLKTLPEN-LSSKVKLQLYSEASVALKLNNPKDFPQELNKOTKMTIDGKELTISP 299
 Db 241 NOLKTLPEN-LSSKVKLQLYSEASVALKLNNPKDFPQELNKOTKMTIDGKELTISP 299
 QY 300 AYLLMDLSAISQSKODEDISASRFEDNEELRSLRSIERHAPWVRNIFVTNGQIPSWLN 359
 Db 300 AYLLMDLSAISQSKODEDISASRFEDNEELRSLRSIERHAPWVRNIFVTNGQIPSWLN 359
 QY 360 LNPRTIYTHODVPRNLSHLPTSSPAIESHRIEGLSQKFTYLNDDVWFGKDWPPDD 419
 Db 360 LNPRTIYTHODVPRNLSHLPTSSPAIESHRIEGLSQKFTYLNDDVWFGKDWPPDD 419
 QY 420 FYSHSKGQKVLTLTPVPCAGCPGSMIKDGYCDKACNNSACDMDGDCSGNSGGRYIA 479
 Db 420 FYSHSKGQKVLTLTPVPCAGCPGSMIKDGYCDKACNNSACDMDGDCSGNSGGRYIA 479
 QY 480 GGGGTSIGVGHPMQFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCFDAGDCGDH 539
 Db 480 GGGGTSIGVGHPMQFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCFDAGDCGDH 539
 QY 540 FHELYVILLPNQTHIIPKGECLPYFSPAELVAKRGEVAGVSDNPIIRHAGIANKKTIH 599
 Db 540 FHELYVILLPNQTHIIPKGECLPYFSPAELVAKRGEVAGVSDNPIIRHAGIANKKTIH 599
 QY 600 LHMHSQNAITTHFNLTFTONTNDEEPKMOITVEVDTRGPKLNSTAQKGENIVSPITLL 659
 Db 600 LHMHSQNAITTHFNLTFTONTNDEEPKMOITVEVDTRGPKLNSTAQKGENIVSPITLL 659
 QY 660 PEABILFEDIPKEXPPKRDVNSTRAOEKVIPLVNTSLBPKDQOLSLNTLDQLE 719
 Db 660 PEABILFEDIPKEXPPKRDVNSTRAOEKVIPLVNTSLBPKDQOLSLNTLDQLE 719
 QY 720 HGDITLKGYNLSKSLALRSFLMNSQHAKIKNOAITTDETNDSLVAPQEKQVHKSILPNSL 779
 Db 720 HGDITLKGYNLSKSLALRSFLMNSQHAKIKNOAITTDETNDSLVAPQEKQVHKSILPNSL 779
 QY 780 GVSESLQRLTPAVSVKNGHOGQONPHLDETTARFRVETHTQKITGANTKEKPSLI 839
 Db 780 GVSESLQRLTPAVSVKNGHOGQONPHLDETTARFRVETHTQKITGANTKEKPSLI 839
 QY 840 VPLESQMTKEKKITGKEKNSRMEENAEHIGTEVLGRKLOHYTDSYLGFLPWEKKKY 899
 Db 840 VPLESQMTKEKKITGKEKNSRMEENAEHIGTEVLGRKLOHYTDSYLGFLPWEKKKY 899
 QY 880 FQDLDDEBSLKTQLAYFTDGRKGRQK 908
 Db 880 FQDLDDEBSLKTQLAYFTDGRKGRQK 908

Search completed: July 26, 2004, 11:09:08
 Job time : 30.6067 secs

Db 361 DNRVATVTHQVDFRNLSHLPTFSSPAISHHRLHRLGLSOKFTYINDVMPFGDWPDPF 420
QY 421 YSHSKQKXYLLTPVENCAGCGGSGSWIKDGYCDKACNNACDMDGDDCGSGSGRYTAG 480
Db 421 YSHSKQKXYLLTPVENCAGCGGSGSWIKDGYCDKACNNACDMDGDDCGSGSGRYTAG 480
QY 481 GGGTGSIGVGHMPQFGGGINSVSYCNQGCANSMADKFCDOACNVLSCGFDAGDGDGQDHF 540
Db 481 GGGTGSIGVGHMPQFGGGINSVSYCNQGCANSMADKFCDOACNVLSCGFDAGDGDGQDHF 540
QY 541 HELYKYLILPNQTHYIIPKGECLPYPSFAEVAKGVEGAYSDNPIIRHASIANKKYTHL 600
Db 541 HELYKYLILPNQTHYIIPKGECLPYPSFAEVAKGVEGAYSDNPIIRHASIANKKYTHL 600
QY 601 IMHSGNNAATTIHNLTFQNTNDEEFMQITVEVDTRBPKLNSTAKGEMLVSPITLLP 660
Db 601 IMHSGNNAATTIHNLTFQNTNDEEFMQITVEVDTRBPKLNSTAKGEMLVSPITLLP 660
QY 661 EAEILFEDIPEKRRPFKRGHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLOEH 720
Db 661 EAEILFEDIPEKRRPFKRGHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLOEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSOAKIKNOAIITDETNDLSVAPOEKVHKSILPNSLG 780
Db 721 GDTTLKGYNLSSKALLRSFLMNSOAKIKNOAIITDETNDLSVAPOEKVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKNGHDGONPDLLETTAFRVEVTHQKTIIGAVTEKRPISLIY 840
Db 781 VSERLQRLTFPAVSVKNGHDGONPDLLETTAFRVEVTHQKTIIGAVTEKRPISLIY 840
QY 841 PLESQWTEKKITGKEKENSMEENAHNIGTEVILGRKLOHYTDSYLGFLPWEKKYF 900
Db 841 PLESQWTEKKITGKEKENSMEENAHNIGTEVILGRKLOHYTDSYLGFLPWEKKYF 900
QY 901 QDLDEESBLKTQLAVFTDSKNTGRQK 928
Db 901 QDLDEESBLKTQLAVFTDSKNTGRQK 928

RESULT 2
US-09-636-077A-1
Sequence 1, Application US/09636077A
Patent No. 6537785
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM
TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
FILE REFERENCE: 195612USO
CURRENT APPLICATION NUMBER: US/09/636,077A
PRIOR FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 928
TYPE: PRF
ORGANISM: Homo sapiens
US-09-636-077A-1

X

Claims to method of treatment

Query Match 100.0%; Score 4907; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRYTTCLSHRVGLVYVCLGVVTVTSAPFQGEVVLWESRDQYHVLPSYDNI 60
Db 1 MLFKLQRYTTCLSHRVGLVYVCLGVVTVTSAPFQGEVVLWESRDQYHVLPSYDNI 60
QY 61 AGSPQNRCLPMPIDVYVTVWNGTDLDELKELQOVEQOMESEKAREILGNTEPTK 120
Db 61 AGSPQNRCLPMPIDVYVTVWNGTDLDELKELQOVEQOMESEKAREILGNTEPTK 120
QY 121 KSEKQCECLTHCIVKPMVLVDPALPANITLKDVPISLPSFHSASDIFNVAKRNPSTNV 180
Db 121 KSEKQCECLTHCIVKPMVLVDPALPANITLKDVPISLPSFHSASDIFNVAKRNPSTNV 180

Db 121 KSEKQCECLTHCIVKPMVLVDPALPANITLKDVPISLPSFHSASDIFNVAKRNPSTNV 180
QY 181 SVVVFSDTKVEDAHSGILKNGSRQTVWNGYLTDTDEKVGVLVYMODLAFSGFPPEFKET 240
Db 181 SVVVFSDTKVEDAHSGILKNGSRQTVWNGYLTDTDEKVGVLVYMODLAFSGFPPEFKET 240
QY 241 NQKTKLPENLSKRYKLLQVSEASVALLKLNPKDFOELNQTQKKNMTIDGKELTISA 300
Db 241 NQKTKLPENLSKRYKLLQVSEASVALLKLNPKDFOELNQTQKKNMTIDGKELTISA 300
QY 301 YLLMDLSAISQSKODDISAREDENEELRYLSRTERAPAWNTFIYTNQIIPSWNL 360
Db 301 YLLMDLSAISQSKODDISAREDENEELRYLSRTERAPAWNTFIYTNQIIPSWNL 360
QY 361 DNRVATVTHQVDFRNLSHLPTFSSPAISHHRLHRLGLSOKFTYINDVMPFGDWPDPF 420
Db 361 DNRVATVTHQVDFRNLSHLPTFSSPAISHHRLHRLGLSOKFTYINDVMPFGDWPDPF 420
QY 421 YSHSKQKXYLLTPVENCAGCGGSGSWIKDGYCDKACNNACDMDGDDCGSGSGRYTAG 480
Db 421 YSHSKQKXYLLTPVENCAGCGGSGSWIKDGYCDKACNNACDMDGDDCGSGSGRYTAG 480
QY 481 GGGTGSIGVGHMPQFGGGINSVSYCNQGCANSMADKFCDOACNVLSCGFDAGDGDGQDHF 540
Db 481 GGGTGSIGVGHMPQFGGGINSVSYCNQGCANSMADKFCDOACNVLSCGFDAGDGDGQDHF 540
QY 541 HELYKYLILPNQTHYIIPKGECLPYPSFAEVAKGVEGAYSDNPIIRHASIANKKYTHL 600
Db 541 HELYKYLILPNQTHYIIPKGECLPYPSFAEVAKGVEGAYSDNPIIRHASIANKKYTHL 600
QY 601 IMHSGNNAATTIHNLTFQNTNDEEFMQITVEVDTRBPKLNSTAKGEMLVSPITLLP 660
Db 601 IMHSGNNAATTIHNLTFQNTNDEEFMQITVEVDTRBPKLNSTAKGEMLVSPITLLP 660
QY 661 EAEILFEDIPEKRRPFKRGHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLOEH 720
Db 661 EAEILFEDIPEKRRPFKRGHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLOEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSOAKIKNOAIITDETNDLSVAPOEKVHKSILPNSLG 780
Db 721 GDTTLKGYNLSSKALLRSFLMNSOAKIKNOAIITDETNDLSVAPOEKVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKNGHDGONPDLLETTAFRVEVTHQKTIIGAVTEKRPISLIY 840
Db 781 VSERLQRLTFPAVSVKNGHDGONPDLLETTAFRVEVTHQKTIIGAVTEKRPISLIY 840
QY 841 PLESQWTEKKITGKEKENSMEENAHNIGTEVILGRKLOHYTDSYLGFLPWEKKYF 900
Db 841 PLESQWTEKKITGKEKENSMEENAHNIGTEVILGRKLOHYTDSYLGFLPWEKKYF 900
QY 901 QDLDEESBLKTQLAVFTDSKNTGRQK 928
Db 901 QDLDEESBLKTQLAVFTDSKNTGRQK 928

RESULT 3
US-09-636-060C-1
Sequence 1, Application US/09636060C
Patent No. 6642038
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM M
TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
FILE REFERENCE: 210119USOCONT
CURRENT APPLICATION NUMBER: US/09/636,060C
PRIOR FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 928
TYPE: PRF
ORGANISM: Homo sapiens

X

Claims to Nucleic acid

US-09-636-060C-1

Query Match 100.0%; Score 4907; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRTYCTLSHRKGLVYCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDYSYRNI 60
DB 1 MLFKLQRTYCTLSHRKGLVYCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDYSYRNI 60
QY 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELKELOQVREOMEEOQKAMEIIGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELKELOQVREOMEEOQKAMEIIGKNTTEPTK 120
QY 121 KSEKQECILTHCIVPMLVLDPALPANITLKDVPSLPSFHSASDIFNVAKPKPNSTNV 180
DB 121 KSEKQECILTHCIVPMLVLDPALPANITLKDVPSLPSFHSASDIFNVAKPKPNSTNV 180
QY 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWRGYLTIDKEVPGVLMQDLAFLSGFPPTFKET 240
DB 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWRGYLTIDKEVPGVLMQDLAFLSGFPPTFKET 240
QY 241 NOLKTKLPENLSKRYLQLYSEASVALLKANNPDPOELNKQTKKMTIDGKELTISPA 300
DB 241 NOLKTKLPENLSKRYLQLYSEASVALLKANNPDPOELNKQTKKMTIDGKELTISPA 300
QY 301 YLLMDLSAISQKODEDISASRFEDNEELRYSLSIRERHAPVVRNIFIVTNGQIPSWMLN 360
DB 301 YLLMDLSAISQKODEDISASRFEDNEELRYSLSIRERHAPVVRNIFIVTNGQIPSWMLN 360
QY 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSOKFIYLLNDVYFGKVPDPDF 420
DB 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSOKFIYLLNDVYFGKVPDPDF 420
QY 421 YSHSGQKXYLLTWVPNCAEGCGPSWIKDGYCDKACNNSACDWDGDCGSGSGSRYIAG 480
DB 421 YSHSGQKXYLLTWVPNCAEGCGPSWIKDGYCDKACNNSACDWDGDCGSGSGSRYIAG 480
QY 481 GGGTGSIGVGHPMQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGCGQDHF 540
DB 481 GGGTGSIGVGHPMQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGCGQDHF 540
QY 541 HELYKYLILPNQTHYIIPKGECLPYFSPAFAVAKRGVAGVSDNPIIRHASIANKWKTILH 600
DB 541 HELYKYLILPNQTHYIIPKGECLPYFSPAFAVAKRGVAGVSDNPIIRHASIANKWKTILH 600
QY 601 IYHSGMNAATTIHFNLTFOVNTDEEFKMOITVEVDIREGPKLNTAKGYENLVSPITLLP 660
DB 601 IYHSGMNAATTIHFNLTFOVNTDEEFKMOITVEVDIREGPKLNTAKGYENLVSPITLLP 660
QY 661 EABLIFEDIPKEKRPFKRHVDNSTRRAOEVEKILPVNISILPQAOULSNTLIDOLEH 720
DB 661 EABLIFEDIPKEKRPFKRHVDNSTRRAOEVEKILPVNISILPQAOULSNTLIDOLEH 720
QY 721 GAITLKGYLSKALLRSFLMNSQAKIKNOAILIDETNDSI.VAFOEQVKSILPNSIG 780
DB 721 GAITLKGYLSKALLRSFLMNSQAKIKNOAILIDETNDSI.VAFOEQVKSILPNSIG 780
QY 781 VSERLQRLTFPAVSXKVGHDGONPDLLETTAFRVEITHOKTIGANTKEKPSLIV 840
DB 781 VSERLQRLTFPAVSXKVGHDGONPDLLETTAFRVEITHOKTIGANTKEKPSLIV 840
QY 841 PLESQMTXKKTITGKESKNSRMEENHIGTEVLARPKLOHYTDSYLGFLPWEKKKTF 900
DB 841 PLESQMTXKKTITGKESKNSRMEENHIGTEVLARPKLOHYTDSYLGFLPWEKKKTF 900
QY 901 QDLDEESLKTOLAYFTDSKNTGRQLK 928
DB 901 QDLDEESLKTOLAYFTDSKNTGRQLK 928

RESULT 4
US-09-986-552-1

; Sequence 1, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIY
; CURRENT APPLICATION NUMBER: US/09/986,552
; PRIOR FILING DATE: 2001-11-09
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-552-1

Query Match 100.0%; Score 4907; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRTYCTLSHRKGLVYCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDYSYRNI 60
DB 1 MLFKLQRTYCTLSHRKGLVYCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDYSYRNI 60
QY 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELKELOQVREOMEEOQKAMEIIGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELKELOQVREOMEEOQKAMEIIGKNTTEPTK 120
QY 121 KSEKQECILTHCIVPMLVLDPALPANITLKDVPSLPSFHSASDIFNVAKPKPNSTNV 180
DB 121 KSEKQECILTHCIVPMLVLDPALPANITLKDVPSLPSFHSASDIFNVAKPKPNSTNV 180
QY 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWRGYLTIDKEVPGVLMQDLAFLSGFPPTFKET 240
DB 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWRGYLTIDKEVPGVLMQDLAFLSGFPPTFKET 240
QY 241 NOLKTKLPENLSKRYLQLYSEASVALLKANNPDPOELNKQTKKMTIDGKELTISPA 300
DB 241 NOLKTKLPENLSKRYLQLYSEASVALLKANNPDPOELNKQTKKMTIDGKELTISPA 300
QY 301 YLLMDLSAISQKODEDISASRFEDNEELRYSLSIRERHAPVVRNIFIVTNGQIPSWMLN 360
DB 301 YLLMDLSAISQKODEDISASRFEDNEELRYSLSIRERHAPVVRNIFIVTNGQIPSWMLN 360
QY 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSOKFIYLLNDVYFGKVPDPDF 420
DB 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSOKFIYLLNDVYFGKVPDPDF 420
QY 421 YSHSGQKXYLLTWVPNCAEGCGPSWIKDGYCDKACNNSACDWDGDCGSGSGSRYIAG 480
DB 421 YSHSGQKXYLLTWVPNCAEGCGPSWIKDGYCDKACNNSACDWDGDCGSGSGSRYIAG 480
QY 481 GGGTGSIGVGHPMQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGCGQDHF 540
DB 481 GGGTGSIGVGHPMQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGCGQDHF 540
QY 541 HELYKYLILPNQTHYIIPKGECLPYFSPAFAVAKRGVAGVSDNPIIRHASIANKWKTILH 600
DB 541 HELYKYLILPNQTHYIIPKGECLPYFSPAFAVAKRGVAGVSDNPIIRHASIANKWKTILH 600
QY 601 IYHSGMNAATTIHFNLTFOVNTDEEFKMOITVEVDIREGPKLNTAKGYENLVSPITLLP 660
DB 601 IYHSGMNAATTIHFNLTFOVNTDEEFKMOITVEVDIREGPKLNTAKGYENLVSPITLLP 660
QY 661 EABLIFEDIPKEKRPFKRHVDNSTRRAOEVEKILPVNISILPQAOULSNTLIDOLEH 720
DB 661 EABLIFEDIPKEKRPFKRHVDNSTRRAOEVEKILPVNISILPQAOULSNTLIDOLEH 720
QY 721 GAITLKGYLSKALLRSFLMNSQAKIKNOAILIDETNDSI.VAFOEQVKSILPNSIG 780

Db 721 GDTILKGNLSALRSFLMSQAKIKNOAILITDEINDSLVAQOEQVKSILPNSLG 780
Qy 781 VSERLQRLTPPAVSVKNGHDGONPPLDLETTARFVETHTOKTIGANTKEKPSLIIV 840
Db 781 VSERLQRLTPPAVSVKNGHDGONPPLDLETTARFVETHTOKTIGANTKEKPSLIIV 840
Qy 841 PLESQMTKEKKTITGKEKENSMEENAHGIVTEVLGRKQHYTDSYLGLPWEKKKYF 900
Db 841 PLESQMTKEKKTITGKEKENSMEENAHGIVTEVLGRKQHYTDSYLGLPWEKKKYF 900
Qy 901 ODLDDEESLKTQALAYFTDSKNTGRQK 928
Db 901 ODLDDEESLKTQALAYFTDSKNTGRQK 928

RESULT 5
US-09-635-872A-15
Sequence 15, Application US/09635872A
Patent No. 6534300
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 195613050
CURRENT APPLICATION NUMBER: US/09/635,872A
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.0
SEQ ID NO 15
LENGTH: 908
TYPE: PRT
ORGANISM: Mus musculus
US-09-635-872A-15

Query Match 79.4%; Score 3894; DB 4; Length 908;
Best Local Similarity 79.9%; Pred. No. 0;
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

Qy 1 MFKLLQRTYCLSHRGVLCFVGAVVTIYSAFOGVEVLEMSRDQHYLFDSDYRNI 60
Db 1 MFKLLQRTYCLSHRGVLCFVGAVVTIYSAFOGVEVLEMSRDQHYLFDSDYRNI 60
Qy 61 AGKSPQNRCLPMPIDVYVTVNGTDLLELKELOQVREMEBEQKAMEIIGKNTTEPTK 120
Db 61 AGKSPQNRCLPMPIDVYVTVNGTDLLELKELOQVREMEBEQKAMEIIGKNTTEPTK 120
Qy 121 KSEKQECILTHCIRKPMVLDPALPANITLKDVSLYPSFHSASDIPNVAKPKNPSTNV 180
Db 121 KSEKQECILTHCIRKPMVLDPALPANITLKDVSLYPSFHSASDIPNVAKPKNPSTNV 180
Qy 181 SVVFPDSTKDVDAHSGLLKGNRSQRTWVGVLTTDKVPGVLVMDLAFISGPPPTFKET 240
Db 181 SVVFPDSTKDVDAHSGLLKGNRSQRTWVGVLTTDKVPGVLVMDLAFISGPPPTFKET 240
Qy 241 NOLKTKLPEN-LSKVKLQLYSEASVALLKLNPKDQOEINKOTKKMTIDGKELTISP 299
Db 241 NOLKTKLPEN-LSKVKLQLYSEASVALLKLNPKDQOEINKOTKKMTIDGKELTISP 299
Qy 300 AYLLWDLAISQSKODEDISASRFEDNEELRYSLSIERHAPVYANFIITVNGOIPSWLN 359
Db 301 AYLLWDLAISQSKODEDISASRFEDNEELRYSLSIERHAPVYANFIITVNGOIPSWLN 360
Qy 360 LDNPRVTIYTHODVVRANISHLPTFSSPAIESHIEHIEGLSQFIYANDVDFGQVWPD 419
Db 361 LDNPRVTIYTHODVVRANISHLPTFSSPAIESHIEHIEGLSQFIYANDVDFGQVWPD 420
Qy 420 FYSHSGQGVYLTVPNCABEGCPGSMIKDGYCDKACNNSACMDVGGDCSGNSGSRVIA 479
Db 421 FYSHSGQGVYLTVPNCABEGCPGSMIKDGYCDKACNNSACMDVGGDCSGNSGSRVIA 480
Qy 480 GGGGTGSGIGVHPWQGGGINSVSYCNGCGANSWLADKFCQACQNVLSGFGDAGDCGDH 539

✓
Claims 21 & 24

Db 481 RGGGTGIGAGHMQGGGGINITISYCNCGANSWLADKFCQACQNVLSGFGDAGDCGDH 540
Qy 540 FHELYVYLLPNQTHYIIPKGECLPYFSAEVAKGVESYSDNPIIRASJANKKTIH 599
Db 541 FHELYVYLLPNQTHYVKEGELYSFSPANIARKIEGYSNDPIIRASJANKKTIH 600
Qy 600 LINESGMATTHFNLTPONTDEEFKQITEVDPREBPKINSTRQKEXYLVSPITL 659
Db 601 LIMPQGNATYITFNLTQNADEEFKQIAEVDTRAPKINSTRQKAYESLVSPITL 660
Qy 660 PEASIIIFEDIPKEKRPPEKRDVNSTRAQSEVKIPLVNISILPBDQASINTLDOLE 719
Db 661 PQADVPEVPEKKEKRPKIRHVDVATGRFOGEVKIPRNIISLPREAVRSLNDIOLE 720
Qy 720 HGDITLKNLSALRSFLMSQAKIKNOAILITDEINDSLVAQOEQVKSILPNSL 779
Db 721 RGDITLKNLSALRSFLMSQAKIKNOAILITDEINDSLVAQOEQVKSILPNSL 777
Qy 780 GVSERLQRLTPPAVSVKNGHDGONPPLDLETTARFVETHTOKTIGANTKEKPSLI 839
Db 781 AGHREBRTAPAEITVYVGRDHALNPVLETNRL-----AQPTLGVTEKENLSPLI 832
Qy 840 VPESQMTKEKKTITGKEKENSMEENAHGIVTEVLGRKQHYTDSYLGLPWEKKKY 899
Db 843 VPESQMTKEKKTITGKEKENSMEENAHGIVTEVLGRKQHYTDSYLGLPWEKKKY 879
Qy 900 FODLDDEESLKTQALAYFTDSKNTGRQK 928
Db 900 FODLDDEESLKTQALAYFTDSKNTGRQK 928

RESULT 6
US-09-636-077A-15
Sequence 15, Application US/09636077A
Patent No. 6537785
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM
TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
FILE REFERENCE: 195612050
CURRENT APPLICATION NUMBER: US/09/636,077A
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.0
SEQ ID NO 15
LENGTH: 908
TYPE: PRT
ORGANISM: Mus musculus
US-09-636-077A-15

Query Match 79.4%; Score 3894; DB 4; Length 908;
Best Local Similarity 79.9%; Pred. No. 0;
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

Qy 1 MFKLLQRTYCLSHRGVLCFVGAVVTIYSAFOGVEVLEMSRDQHYLFDSDYRNI 60
Db 1 MFKLLQRTYCLSHRGVLCFVGAVVTIYSAFOGVEVLEMSRDQHYLFDSDYRNI 60
Qy 61 AGKSPQNRCLPMPIDVYVTVNGTDLLELKELOQVREMEBEQKAMEIIGKNTTEPTK 120
Db 61 AGKSPQNRCLPMPIDVYVTVNGTDLLELKELOQVREMEBEQKAMEIIGKNTTEPTK 120
Qy 121 KSEKQECILTHCIRKPMVLDPALPANITLKDVSLYPSFHSASDIPNVAKPKNPSTNV 180
Db 121 KSEKQECILTHCIRKPMVLDPALPANITLKDVSLYPSFHSASDIPNVAKPKNPSTNV 180
Qy 181 SVVFPDSTKDVDAHSGLLKGNRSQRTWVGVLTTDKVPGVLVMDLAFISGPPPTFKET 240
Db 181 SVVFPDSTKDVDAHSGLLKGNRSQRTWVGVLTTDKVPGVLVMDLAFISGPPPTFKET 240
Qy 241 NOLKTKLPEN-LSKVKLQLYSEASVALLKLNPKDQOEINKOTKKMTIDGKELTISP 299

Db 241 SGLTKTLPRKAFPLKIKTLKTLKLYSEASVALLKLNKPPGQELNNQTKKNTITDKGELLTSP 300

Qy 300 AYLLMDLSAISQSDQEDDISASRPEDNEELRYSLSRISERHAPVNRNIFIVTNGQIPSWLN 359

Db 301 AYLLMDLSAISQSDQEDDISASRPEDNEELRYSLSRISERHAPVNRNIFIVTNGQIPSWLN 360

Qy 360 LDNRVATIVHODVFRNLSHLPTFSSPALESIHRIEGLSOKFIYLANDVMFGKVPDPD 419

Db 361 LDNRVATIVHODVFRNLSHLPTFSSPALESIHRIEGLSOKFIYLANDVMFGKVPDPD 420

Qy 420 FYHSKQKQKYYLLTWPVNCACGCPGSKYIKQYCDKACNNSACDWDGDCSGSGSRYYLA 479

Db 421 FYHSKQKQKYYLLTWPVNCACGCPGSKYIKQYCDKACNNSPCMDGNGCSGTAGNRFFVA 480

Qy 480 GGGGTGSIIGVGHWPQFGGINSVSYCMQGCANSMLADKFCQDQACNVLSGCPADPGQGDH 539

Db 481 RGGGTGSIIGVGHWPQFGGINSVSYCMQGCANSMLADKFCQDQACNVLSGCPADPGQGDH 540

Qy 540 FHELYKTLIPNQHYYIIPKGECLPIPSFAEVAKGVGAYSDNPIIRHASTANKKTIH 599

Db 541 FHELYKTLIPNQHYYIIPKGECLPIPSFAEVAKGVGAYSDNPIIRHASTANKKTIH 600

Qy 600 LIMSANNATIIHNLIFONTNDEEFMQITVEVDREGKLNSTAKQKXYNLSPITLL 659

Db 601 LIMSANNATIIHNLIFONTNDEEFMQITVEVDREGKLNSTAKQKXYNLSPITLL 660

Qy 660 PEALIEEDIPKPKRPFKFRHDVNSTRPAQEEVKIPLVINISLPPQAGLSINTLDLQLE 719

Db 661 PQADVPEDVPEKPKRPFKIRHDVNSTRPAQEEVKIPLVINISLPPQAGLSINTLDLQLE 720

Qy 720 HGDITLGYNLSKSLALRSFLMNSQAKIINGQAIITDETNDLSVVAQECQVHKSLLPNSL 779

Db 721 RGDITLGYNLSKSLALRSFLMNSQAKIINGQAIITDETNDLSVVAQECQVHKSLLPNSL 777

Qy 780 GVSERLQRLTFPAVSVKVNGHDOGNPDLDETTAARFVETHQTKTIGENVTKKPPSLI 839

Db 778 AGEHRSERMTAPAEIVTVKGRDHALNPPVLETNARL-----AQPGLVTVSEKENSPLI 832

Qy 840 VPLSEONTKEKLTGKEKENSMBEENAKENHIGTEVLGRKLOHYHDSYLSGLPWEKKY 899

Db 833 VPPSHLP-----KEESDRABGNA--VPKXELVPGRRLO--QNTPGFLPWEKKY 879

Qy 900 FQDILDEESIKTQLAYFTDSKNTGRQLK 928

Db 880 FQDILDEESIKTQLAYFTDRKHTRQLK 908

RESULT 7

US-09-636-060C-15

; Sequence 15, Application US/09636060C

; Patent No. 6642038

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM M

; TITLE OF INVENTION: GLCNAIC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY

; FILE REFERENCE: 210119USOCONT

; CURRENT APPLICATION NUMBER: US/09/636,060C

; CURRENT FILING DATE: 2000-08-10

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: SeqID version 3.1

; SEQ ID NO 15

; LENGTH: 908

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-636-060C-15

Query Match	79.4%	Score 3894	DB 4	Length 908
Best local Similarity	79.9%	Pred. No. 0		
Matches 742	Conservative 61	Mismatches 104	Indels 22	Gaps 7

```

Db      1 MLFLLKQRYTLCLSHRGVLVCYFVGVVVTIVSAQFQEBVVLWENSRDQYHVLFDYSRNDI 60
QY      61 AGKSFONFLCLPMP-IDVYVTVNNGTDELLKELQVREQMEBEOKAMREILCKNTTEPTK 120
Db      61 AGKSFONFLCLPMP-IDVYVTVNNGTDELLKELQVREHMEBEOKAMRETLCKNTTEPTK 120
QY      121 KSEKQJECLELPHCFKVMPVLVDPALPANIITLKDVSLKPSFASADIFNVAKPKPSTNV 180
Db      121 KSEKQJECLELPHCFKVMPVLVDPALPANIITLKDVSLKPSFASADIFNVAKPKPSTNV 180
QY      181 SVVVFDS-TKVDYEDAHSGILLKGNRSQTVWRYGLTTDKVEYPGVLVMQDLAFLSGFPPTFKET 240
Db      181 PVVVFEDTLKQVEDAHAGFPKGGQQTQDVWRAVILTTDKDAPGLVILIGLAFLSGFPPTFKET 240
QY      241 NQKTKLPEN-LSSKVKLLQLYSEASVALLKLNPKDQELNKQTKKMMTIDGKLLTSP 299
Db      241 NQKTKLPEN-LSSKVKLLQLYSEASVALLKLNPKDQELNKQTKKMMTIDGKLLTSP 299
QY      300 AYLLMDLSAISQSKODEDIDISASRFEDNEELRYLSRIERHAPWVWNIETVINGQJPSMLN 359
Db      301 AYLLMDLSAISQSKODEDIDISASRFEDNEELRYLSRIERHAPWVWNIETVINGQJPSMLN 360
QY      360 LDNRPVTVTVTHQVDFRNLSHLPTFSSPAIESHIERIEGLSQXYTLNDQVDFGKDVWPD 419
Db      361 LDNRPVTVTVTHQVDFRNLSHLPTFSSPAIESHIERIEGLSQXYTLNDQVDFGKDVWPD 420
QY      420 FYSHSKGXVVLTVMPVPCAGCGPGSWTKDQYCDKACNNASADMGGCGSGSGSRYIA 479
Db      421 FYSHSKGXVVLTVMPVPCAGCGPGSWTKDQYCDKACNNASADMGGCGSGSGSRYIA 480
QY      480 GGGGTGSIQVGHPMQFGGGINSVSYCNQGCANSWLADKFCQDQACNVLSGCFPAGCGQDH 539
Db      481 RGGGTGNIQAQGHMQFGGGINTVISYCNQGCANSWLADKFCQDQACNVLSGCFPAGCGQDH 540
QY      540 FHELYKVLLENVQHYIIPKGECLPYESFAEYAKRGVAGANSDNPIIRHASJANKKXTIH 599
Db      541 FHELYKVLLENVQHYIIPKGECLPYESFAEYAKRGVAGANSDNPIIRHASJANKKXTIH 600
QY      600 LHMGGMAATIIHENLTFONTNDEBPKQITVEVDTREGPKLNSTPAOKGYENLVSPITLL 659
Db      601 LHMGGMAATIIHENLTFONTNDEBPKQITVEVDTREGPKLNSTPAOKGYENLVSPITLL 660
QY      660 PEAEILFEDIPKEXRPFKRDVNSTRAOEBVKIPLVNISLBPKAQSLNTLDLOIE 719
Db      661 POADVPEDVPEKEXRPFKIRRDVNAVATGRFOEBVKIPLVNISLBPKEKAQVRLSNDLOIE 720
QY      720 HGDITLKKNYLSKALLSPLMNSQHAKIKNOAIIITDENTNSIVAPQEKQVYSKILPNSL 779
Db      721 RGDITLKKNYLSKALLSPLMNSQHAKIKNOAIIITDENTNSIVAPQEKQVYSKILPNSL 779
QY      780 GVSERLQGLTFPAASVKVNGHDOGNPDLDETTFARFVETHTQKTIIGANVYKEKPSLI 839
Db      778 AGEHRSSEWTPAELVTVYKGRDHLNPPVLETNARL-----AQPTLGTVYSKEMLSPLI 832
QY      840 VPLSQMTKEKKITGKEKENSRMENAEENHIGYTEVLLGRKLOHYTDSYLGLPWEKKKY 899
Db      833 VPESHLP-----KEEESDRABGNA--VPVKELVGRRLQ--QNYPGFLPWEKKKY 879
QY      900 FODLLDEESLTKQALAYFTDSKNTGRQLK 928
Db      880 FODLLDEESLTKQALAYFTDRKNTGRQLK 908

```

```

RESULT 8
US-09-986-552-15
; Sequence 15, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 2150890577DIV
; CURRENT APPLICATION NUMBER: US/09/986,552

```

; CURRENT FILING DATE: 2001-11-09
 ; PRIOR APPLICATION NUMBER: 09/635,872
 ; PRIOR FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; PRIOR FILING DATE: 1999-09-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 15
 ; LENGTH: 908
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-09-986-552-15

Query Match 79.4%; Score 3894; DB 4; Length 908;
 Best Local Similarity 79.9%; Pred. No. 0;
 Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MFKLLQROTTCLSHRGVLVCPGVVTVTISAFQFGVVLVEMSDQYHLFDSYRNI 60
 DB 1 MFKLLQROTTCLSHRGVLVCPGVVTVTISAFQFGVVLVEMSDQYHLFDSYRNI 60
 QY 61 AGSFQNRCLPMPIDVYVWNGTDLBELKELQVREMEBEQAMEELGKTEPTK 120
 DB 61 AGSFQNRCLPMPIDVYVWNGTDLBELKELQVREMEBEQAMEELGKTEPTK 120
 QY 121 KSEKQECLETHCIVKPMVLDPALPANTLKDVPSLSPSBSADIFVAKKPKPSTNV 180
 DB 121 KSEKQECLETHCIVKPMVLDPALPANTLKDVPSLSPSBSADIFVAKKPKPSTNV 180
 QY 181 SYVVPDSTDVEDASGLKGNRSQTVWRGVLTDTKEVPGVLVMDLAFLESGPPTFKET 240
 DB 181 SYVVPDSTDVEDASGLKGNRSQTVWRGVLTDTKEVPGVLVMDLAFLESGPPTFKET 240
 QY 241 NQKTLQTPEN-LSKVKLIQLYSEASVALLKLNPKDQOELNKQTKKMTIDGKELTISP 299
 DB 241 NQKTLQTPEN-LSKVKLIQLYSEASVALLKLNPKDQOELNKQTKKMTIDGKELTISP 299
 QY 300 ATLLMDLSAISQSKODEDASASRFDNEELRYSLSIRSHAPWVNIPIVTNGQIPSWLN 359
 DB 301 ATLLMDLSAISQSKODEDASASRFDNEELRYSLSIRSHAPWVNIPIVTNGQIPSWLN 360
 QY 360 LNPRTVITVHODVPRNLSHLPFSSPAIESHIREGLSOKETIYLANDVYFGDVPDD 419
 DB 361 LNPRTVITVHODVPRNLSHLPFSSPAIESHIREGLSOKETIYLANDVYFGDVPDD 420
 QY 420 FYSHSGQKAVLTWPVPCNCAEGCPGSMWKDGYCDKACNNSACMDGDPGSGNSGSRYA 479
 DB 421 FYSHSGQKAVLTWPVPCNCAEGCPGSMWKDGYCDKACNNSACMDGDPGSGNSGSRYA 480
 QY 480 GGGGTGSGVGHWPQFGGINSVSYCNOGCANSWLADFCDOACNVLSCGFDAGDCGDH 539
 DB 481 RGGGTGSGVGHWPQFGGINSVSYCNOGCANSWLADFCDOACNVLSCGFDAGDCGDH 540
 QY 540 FHEIKVTLIPNQTHTIIPKGECLPYSPAPYAKGVGASDNPITTHASITANKMTIH 599
 DB 541 FHEIKVTLIPNQTHTIIPKGECLPYSPAPYAKGVGASDNPITTHASITANKMTIH 600
 QY 600 LTHSGMNAATTIHFNLTFONTNDEEFKQIIVEDVTRBGPKLNTAQKGVENLVSPTLL 659
 DB 601 LTHSGMNAATTIHFNLTFONTNDEEFKQIIVEDVTRBGPKLNTAQKGVENLVSPTLL 660
 QY 660 PBAELIFEDIPEKKEFPKPKRDVNSTRAQGEVKIPLVNTSLPKDQLSLNTLDQLE 719
 DB 661 PBAELIFEDIPEKKEFPKPKRDVNSTRAQGEVKIPLVNTSLPKDQLSLNTLDQLE 720
 QY 720 HODITLKGNLSKSLNSFLNNSOHAQIKNOAITDDETNSLVAPQKQVHKSLIPNSL 779
 DB 721 RQDITLKGNLSKSLNSFLNNSOHAQIKNOAITDDETNSLVAPQKQVHKSLIPNSL 777
 QY 780 GVSESLQRLTPPAVSVKNGHDQGNPPLDLETTARKFVEVTHYTKTGNTKKEKPSLI 839
 DB 778 AGEHSRSEWTAFAETVYKGRDHALNPPVLETNARL-----AOPFLGVTVSKENSLPLI 832

QY 840 VPLESQTKKKTGKEKENSMEENHIGTVEVLGRKLOHYTDSYLGFLPWKCKY 899
 DB 833 VPESHLP-----KEESDRAGNA---VPKELVPRRIQ---QNPGLPWKCKY 879
 QY 900 FQDLDBEESTKQIAYFTDSKTRGRQK 928
 DB 880 FQDLDBEESTKQIAYFTDRKHTGRQK 908

RESULT 9

US-09-635-872A-11
 ; Sequence 11; Application US/09635872A
 ; Patent No. 6534300
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
 ; FILE REFERENCE: 195612050
 ; CURRENT APPLICATION NUMBER: US/09/635,872A
 ; CURRENT FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; PRIOR FILING DATE: 1999-09-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 11
 ; LENGTH: 113
 ; TYPE: PRT
 ; ORGANISM: Rattus rattus
 ; US-09-635-872A-11

Query Match 10.3%; Score 503; DB 4; Length 113;
 Best Local Similarity 92.5%; Pred. No. 1,1e-36;
 Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 233 FPPFTKENTQKTKLPENSSKVKLIQLYSEASVALLKLNPKDQOELNKQTKKMTIDG 292
 DB 1 FPPFTKENTQKTKLPENSSKVKLIQLYSEASVALLKLNPKDQOELNKQTKKMTIDG 60
 QY 293 KELTISPAYLMDLSAISQSKODEDASASRFDNEELRYSLSIRSH 339
 DB 61 KELTISPAYLMDLSAISQSKODEDASASRFDNEELRYSLSIRSH 107

RESULT 10

US-09-636-077A-11
 ; Sequence 11; Application US/09636077A
 ; Patent No. 6537765
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM
 ; TITLE OF INVENTION: METHODS OF TREATING LYOSOMAL STORAGE DISEASE
 ; FILE REFERENCE: 195612050
 ; CURRENT APPLICATION NUMBER: US/09/636,077A
 ; CURRENT FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; PRIOR FILING DATE: 1999-09-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 11
 ; LENGTH: 113
 ; TYPE: PRT
 ; ORGANISM: Rattus rattus
 ; US-09-636-077A-11

Query Match 10.3%; Score 503; DB 4; Length 113;
 Best Local Similarity 92.5%; Pred. No. 1,1e-36;
 Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 233 FPPFTKENTQKTKLPENSSKVKLIQLYSEASVALLKLNPKDQOELNKQTKKMTIDG 292
 DB 1 FPPFTKENTQKTKLPENSSKVKLIQLYSEASVALLKLNPKDQOELNKQTKKMTIDG 60
 QY 293 KELTISPAYLMDLSAISQSKODEDASASRFDNEELRYSLSIRSH 339
 DB 61 KELTISPAYLMDLSAISQSKODEDASASRFDNEELRYSLSIRSH 107

RESULT 11
US-09-636-060C-11
; Sequence 11, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLUTAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US00CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 11
; LENGTH: 113
; TYPE: PRF
; ORGANISM: Rattus rattus
US-09-636-060C-11

Query Match 10.3%; Score 503; DB 4; Length 113;
Best Local Similarity 92.5%; Pred. No. 1.1e-36;
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 233 PEPFRETQKTKLPENLSSKVKLQLYSEASVALTLKNNPKDQELNKKQTKMTIDG 292
DB 1 PEPFRETQKTKLPENLSSKVKLQLYSEASVALTLKNNPKDQELNKKQTKMTIDG 60
QY 293 KELTISPAYLWDLAISQSKODEDISASRFEDNEELRYSLSIERH 339
DB 61 KELTISPAYLWDLAISQSKODEDVASASRFEDNEELRYSLSIERH 107

RESULT 12
US-09-986-552-11
; Sequence 11, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 21089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 11
; LENGTH: 113
; TYPE: PRF
; ORGANISM: Rattus rattus
US-09-986-552-11

Query Match 10.3%; Score 503; DB 4; Length 113;
Best Local Similarity 92.5%; Pred. No. 1.1e-36;
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 233 PEPFRETQKTKLPENLSSKVKLQLYSEASVALTLKNNPKDQELNKKQTKMTIDG 292
DB 1 PEPFRETQKTKLPENLSSKVKLQLYSEASVALTLKNNPKDQELNKKQTKMTIDG 60
QY 293 KELTISPAYLWDLAISQSKODEDISASRFEDNEELRYSLSIERH 339
DB 61 KELTISPAYLWDLAISQSKODEDVASASRFEDNEELRYSLSIERH 107

RESULT 13
US-09-635-872A-13
; Sequence 13, Application US/09635872A

; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 13
; LENGTH: 502
; TYPE: PRF
; ORGANISM: Drosophila melanogaster
US-09-635-872A-13

Query Match 9.8%; Score 479; DB 4; Length 502;
Best Local Similarity 55.4%; Pred. No. 1.8e-33;
Matches 82; Conservative 29; Mismatches 37; Indels 0; Gaps 0;

QY 322 RFEDNEELRYSLSIERHAPWBNFIYVINGQIPSWLNDNPRVTIVTHQDVFRLNSHP 381
DB 4 RFDDXNELRYSLSIERHAPWBNFIYVINGQIPSWLNDNPRVTIVTHQDVFRLNSHP 63
QY 382 TFSSPAIESHIRIEGISOKEIYLANDVMPGKDWDPDYSKGGKXYLTWVPVPCAG 441
DB 64 TFSSPAIESHIRIEGISOKEIYLANDVMPGKDWDPDYSKGGKXYLTWVPVPCAG 123
QY 442 CPGSWKDGYCDKACNNSACDMDGDCS 469
DB 124 CPWTYIGDAGCDRHCNIDACQPDGDCS 151

RESULT 14
US-09-636-077A-13
; Sequence 13, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 13
; LENGTH: 502
; TYPE: PRF
; ORGANISM: Drosophila melanogaster
US-09-636-077A-13

Query Match 9.8%; Score 479; DB 4; Length 502;
Best Local Similarity 55.4%; Pred. No. 1.8e-33;
Matches 82; Conservative 29; Mismatches 37; Indels 0; Gaps 0;

QY 322 RFEDNEELRYSLSIERHAPWBNFIYVINGQIPSWLNDNPRVTIVTHQDVFRLNSHP 381
DB 4 RFDDXNELRYSLSIERHAPWBNFIYVINGQIPSWLNDNPRVTIVTHQDVFRLNSHP 63
QY 382 TFSSPAIESHIRIEGISOKEIYLANDVMPGKDWDPDYSKGGKXYLTWVPVPCAG 441
DB 64 TFSSPAIESHIRIEGISOKEIYLANDVMPGKDWDPDYSKGGKXYLTWVPVPCAG 123
QY 442 CPGSWKDGYCDKACNNSACDMDGDCS 469
DB 124 CPWTYIGDAGCDRHCNIDACQPDGDCS 151

RESULT 15
US-09-636-060C-13

QY 181 SVVFPDSTKDVEDAHSGLLKGNRSQTVWVGVLTTDKVPGVLVLMODLAFLSGFPPTFKET 240
DB 181 SVVFPDSTKDVEDAHSGLLKGNRSQTVWVGVLTTDKVPGVLVLMODLAFLSGFPPTFKET 240
QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDPFOLNKQTKKNTIDGKELTISPA 300
DB 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDPFOLNKQTKKNTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLSIESHAPVWRNIFVTNGQILPSWML 360
DB 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLSIESHAPVWRNIFVTNGQILPSWML 360
QY 361 DNPRTVIVHODVFNRLSHLPFSSPAIESHIREGLSQKFIYLANDVMFGKVPDPDF 420
DB 361 DNPRTVIVHODVFNRLSHLPFSSPAIESHIREGLSQKFIYLANDVMFGKVPDPDF 420
QY 421 YHSHKQKQKYLITWPNCAEGCPGSMIKDGYCDKACNNASACDMDGDCSGNSGSRITAG 480
DB 421 YHSHKQKQKYLITWPNCAEGCPGSMIKDGYCDKACNNASACDMDGDCSGNSGSRITAG 480
QY 481 GGGTGSIGVGHFWOGGGINSVSYCNOCGANSMLADKFCDDACNTLSCGPDAGDGDHF 540
DB 481 GGGTGSIGVGHFWOGGGINSVSYCNOCGANSMLADKFCDDACNTLSCGPDAGDGDHF 540
QY 541 HELYKYLILPNOTHYIIPKGECLPYFSFAEVAKRGVEGAYSNDPIIRHASIANKWKTIHL 600
DB 541 HELYKYLILPNOTHYIIPKGECLPYFSFAEVAKRGVEGAYSNDPIIRHASIANKWKTIHL 600
QY 601 IAHSGNNAATTIHFNLTFOQNTDEEFKMOITVEVDTRREGPKLNSTAKQKYNLVSPITLLP 660
DB 601 IAHSGNNAATTIHFNLTFOQNTDEEFKMOITVEVDTRREGPKLNSTAKQKYNLVSPITLLP 660
QY 661 EAEIILFEDIPEKRRPKFRHDVNSTRAQEEVKIPLVNISSLPPDQQLSINTLDLOEH 720
DB 661 EAEIILFEDIPEKRRPKFRHDVNSTRAQEEVKIPLVNISSLPPDQQLSINTLDLOEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIIIDETNDSLVAPQEKQVHKSILPNSLG 780
DB 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIIIDETNDSLVAPQEKQVHKSILPNSLG 780
QY 781 VSEIRLQRLTFPAVSXKVNKHOGQNPDLDETTAFRVEITHQKTIIGANTKEKPPSLIV 840
DB 781 VSEIRLQRLTFPAVSXKVNKHOGQNPDLDETTAFRVEITHQKTIIGANTKEKPPSLIV 840
QY 841 PLESQWTEKKITGKEKENSMEENAEENHIGTEVILGRKLOHYTDSYLGFLPWKCKKYF 900
DB 841 PLESQWTEKKITGKEKENSMEENAEENHIGTEVILGRKLOHYTDSYLGFLPWKCKKYF 900
QY 901 QDLDBEESLKTQLAYFTDSKNTGRQK 928
DB 901 QDLDBEESLKTQLAYFTDSKNTGRQK 928

RESULT 2

US-09-986-552-1
Sequence 1, Application US/09986552
Patent No. US20020150981A1
GENERAL INFORMATION:
APPLICANT: CANFIELD, William
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 215083US7DIIV
CURRENT APPLICATION NUMBER: US/09/986,552
CURRENT FILING DATE: 2001-11-09
PRIOR FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 928
TYPE: PRF
ORGANISM: Homo sapiens

US-09-986-552-1
Query Match 100.0%; Score 4907; DB 9; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEFKLIOROTYTCLSRYGLYCYFLGVVTTYISARQFGVVLNERSRDQYHVLFDSDYRDI 60
DB 1 MEFKLIOROTYTCLSRYGLYCYFLGVVTTYISARQFGVVLNERSRDQYHVLFDSDYRDI 60
QY 61 AGSFQNRCLPMPIDVYVTVWNGTDLLELKELOQREBMEBEQKAMEIILKNTTEPTK 120
DB 61 AGSFQNRCLPMPIDVYVTVWNGTDLLELKELOQREBMEBEQKAMEIILKNTTEPTK 120
QY 121 KSEKQLECLTHICIKYPMVLVDPALPANTLTKDVSILPSFHSASDIPNNAKPNSTNV 180
DB 121 KSEKQLECLTHICIKYPMVLVDPALPANTLTKDVSILPSFHSASDIPNNAKPNSTNV 180
QY 181 SVVFPDSTKDVEDAHSGLLKGNRSQTVWVGVLTTDKVPGVLVLMODLAFLSGFPPTFKET 240
DB 181 SVVFPDSTKDVEDAHSGLLKGNRSQTVWVGVLTTDKVPGVLVLMODLAFLSGFPPTFKET 240
QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDPFOLNKQTKKNTIDGKELTISPA 300
DB 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDPFOLNKQTKKNTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLSIESHAPVWRNIFVTNGQILPSWML 360
DB 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLSIESHAPVWRNIFVTNGQILPSWML 360
QY 361 DNPRTVIVHODVFNRLSHLPFSSPAIESHIREGLSQKFIYLANDVMFGKVPDPDF 420
DB 361 DNPRTVIVHODVFNRLSHLPFSSPAIESHIREGLSQKFIYLANDVMFGKVPDPDF 420
QY 421 YHSHKQKQKYLITWPNCAEGCPGSMIKDGYCDKACNNASACDMDGDCSGNSGSRITAG 480
DB 421 YHSHKQKQKYLITWPNCAEGCPGSMIKDGYCDKACNNASACDMDGDCSGNSGSRITAG 480
QY 481 GGGTGSIGVGHFWOGGGINSVSYCNOCGANSMLADKFCDDACNTLSCGPDAGDGDHF 540
DB 481 GGGTGSIGVGHFWOGGGINSVSYCNOCGANSMLADKFCDDACNTLSCGPDAGDGDHF 540
QY 541 HELYKYLILPNOTHYIIPKGECLPYFSFAEVAKRGVEGAYSNDPIIRHASIANKWKTIHL 600
DB 541 HELYKYLILPNOTHYIIPKGECLPYFSFAEVAKRGVEGAYSNDPIIRHASIANKWKTIHL 600
QY 601 IAHSGNNAATTIHFNLTFOQNTDEEFKMOITVEVDTRREGPKLNSTAKQKYNLVSPITLLP 660
DB 601 IAHSGNNAATTIHFNLTFOQNTDEEFKMOITVEVDTRREGPKLNSTAKQKYNLVSPITLLP 660
QY 661 EAEIILFEDIPEKRRPKFRHDVNSTRAQEEVKIPLVNISSLPPDQQLSINTLDLOEH 720
DB 661 EAEIILFEDIPEKRRPKFRHDVNSTRAQEEVKIPLVNISSLPPDQQLSINTLDLOEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIIIDETNDSLVAPQEKQVHKSILPNSLG 780
DB 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIIIDETNDSLVAPQEKQVHKSILPNSLG 780
QY 781 VSEIRLQRLTFPAVSXKVNKHOGQNPDLDETTAFRVEITHQKTIIGANTKEKPPSLIV 840
DB 781 VSEIRLQRLTFPAVSXKVNKHOGQNPDLDETTAFRVEITHQKTIIGANTKEKPPSLIV 840
QY 841 PLESQWTEKKITGKEKENSMEENAEENHIGTEVILGRKLOHYTDSYLGFLPWKCKKYF 900
DB 841 PLESQWTEKKITGKEKENSMEENAEENHIGTEVILGRKLOHYTDSYLGFLPWKCKKYF 900
QY 901 QDLDBEESLKTQLAYFTDSKNTGRQK 928
DB 901 QDLDBEESLKTQLAYFTDSKNTGRQK 928

RESULT 3
US-10-023-888-4

```
/ Sequence 4, Application US/10023888
/ Publication No. US20030119088A1
/ GENERAL INFORMATION:
/ APPLICANT: CANFIELD, William
/ TITLE OF INVENTION: SOLUBLE GLUCNAC PHOSPHOTRANSFERASE
/ FILE REFERENCE: 203515US77
/ CURRENT APPLICATION NUMBER: US/10/023, 888
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Patent version 3.1
/ SEQ ID NO 4
/ LENGTH: 928
/ TYPE: PRF
/ ORGANISM: Homo sapiens
US-10-023-888-4
```

```
Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MFKLQROTYTCLSHRYGLVYCFGLGVVYTVISAFQFGEVYVLEMSRDQYHVLFDSDYRDN 60
DB 1 MFKLQROTYTCLSHRYGLVYCFGLGVVYTVISAFQFGEVYVLEMSRDQYHVLFDSDYRDN 60
QY 61 AGKSFQNRCLCLPMPIDVYTVWNGTDLLELKELOQVREQMEBEQKAMREILGKNTTEPTK 120
DB 61 AGKSFQNRCLCLPMPIDVYTVWNGTDLLELKELOQVREQMEBEQKAMREILGKNTTEPTK 120
QY 121 KSEKQLECLTHCIKYPMLVLDPALPANITLKDVPSLYPSFSASDIPVAKPKNSTNV 180
DB 121 KSEKQLECLTHCIKYPMLVLDPALPANITLKDVPSLYPSFSASDIPVAKPKNSTNV 180
QY 181 SYVVFSTKDVDAHSGLLKGNRSQTVWNGYLTDTKEVGLVLMODLAFISGFPPTFKET 240
DB 181 SYVVFSTKDVDAHSGLLKGNRSQTVWNGYLTDTKEVGLVLMODLAFISGFPPTFKET 240
QY 241 NQKTKLPENLSSKVKLLQLYSEASVALKLNPKDFOELNKQTKKMTIDGKELTISPA 300
DB 241 NQKTKLPENLSSKVKLLQLYSEASVALKLNPKDFOELNKQTKKMTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLRSIERHAPVWRNIPIYVNGQIPSWNL 360
DB 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLRSIERHAPVWRNIPIYVNGQIPSWNL 360
QY 361 DNPRTVITVHQDVFNRLSHLPTSSPAIESHIRIEGLSQKFTYLNDDVMFGKDWPPDF 420
DB 361 DNPRTVITVHQDVFNRLSHLPTSSPAIESHIRIEGLSQKFTYLNDDVMFGKDWPPDF 420
QY 421 YSHSGQKQYVLTWVPVNCAGCGPGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYTAG 480
DB 421 YSHSGQKQYVLTWVPVNCAGCGPGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYTAG 480
QY 481 GGGTGSIGVGHPMQFGGGINSVYCNQGCANSMADKFCDOACNVLSGCFDAGDCQDHF 540
DB 481 GGGTGSIGVGHPMQFGGGINSVYCNQGCANSMADKFCDOACNVLSGCFDAGDCQDHF 540
QY 541 HELYKYLIPNQTHYIIPKGECLPYPSFAEVAKRGVGAYSNDPIIRHASIANKKTIHL 600
DB 541 HELYKYLIPNQTHYIIPKGECLPYPSFAEVAKRGVGAYSNDPIIRHASIANKKTIHL 600
QY 601 IHSGMNATTIHFNLTFOVNDDEFPMQITVEVDVREBPKLNSTAQKYEVLVSPITLLP 660
DB 601 IHSGMNATTIHFNLTFOVNDDEFPMQITVEVDVREBPKLNSTAQKYEVLVSPITLLP 660
QY 661 EAEIIPEDIPEKREPEKFGADVSTRQAQEVKIPLVNISLLPRDAQLSANTLDLQIEH 720
DB 661 EAEIIPEDIPEKREPEKFGADVSTRQAQEVKIPLVNISLLPRDAQLSANTLDLQIEH 720
QY 721 GQITLKGYNLSKALLRSFLMNSOHAQIKNOAILIDENDNDSLVAQEOGVKSTILPNSIG 780
DB 721 GQITLKGYNLSKALLRSFLMNSOHAQIKNOAILIDENDNDSLVAQEOGVKSTILPNSIG 780
QY 781 VSERLQRLTPEAVSVKVNHGQGNPDLLETTAARFVETHQTKTIGGNVTKKEKPSLIV 840
```

```
DB 781 VSERLQRLTPEAVSVKVNHGQGNPDLLETTAARFVETHQTKTIGGNVTKKEKPSLIV 840
QY 841 PLESQMTKEKKTIGKEKENSMEENAHNIGVTEVLLGRDLQHYTDSYGLFPEWKKKYF 900
DB 841 PLESQMTKEKKTIGKEKENSMEENAHNIGVTEVLLGRDLQHYTDSYGLFPEWKKKYF 900
QY 901 QDLIDEEESLKTQLAFTDSKNTGRQLK 928
DB 901 QDLIDEEESLKTQLAFTDSKNTGRQLK 928
```

```
RESULT 4
US-10-023-889-4
/ Sequence 4, Application US/10023889
/ Publication No. US20030124652A1
/ GENERAL INFORMATION:
/ APPLICANT: CANFIELD, William
/ TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARB
```

```
/ FILE REFERENCE: 203512US77
/ CURRENT APPLICATION NUMBER: US/10/023, 889
/ NUMBER OF SEQ ID NOS: 21
/ SOFTWARE: Patent version 3.1
/ SEQ ID NO 4
/ LENGTH: 928
/ TYPE: PRF
/ ORGANISM: Homo sapiens
US-10-023-889-4
```

```
Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MFKLQROTYTCLSHRYGLVYCFGLGVVYTVISAFQFGEVYVLEMSRDQYHVLFDSDYRDN 60
DB 1 MFKLQROTYTCLSHRYGLVYCFGLGVVYTVISAFQFGEVYVLEMSRDQYHVLFDSDYRDN 60
QY 61 AGKSFQNRCLCLPMPIDVYTVWNGTDLLELKELOQVREQMEBEQKAMREILGKNTTEPTK 120
DB 61 AGKSFQNRCLCLPMPIDVYTVWNGTDLLELKELOQVREQMEBEQKAMREILGKNTTEPTK 120
QY 121 KSEKQLECLTHCIKYPMLVLDPALPANITLKDVPSLYPSFSASDIPVAKPKNSTNV 180
DB 121 KSEKQLECLTHCIKYPMLVLDPALPANITLKDVPSLYPSFSASDIPVAKPKNSTNV 180
QY 181 SYVVFSTKDVDAHSGLLKGNRSQTVWNGYLTDTKEVGLVLMODLAFISGFPPTFKET 240
DB 181 SYVVFSTKDVDAHSGLLKGNRSQTVWNGYLTDTKEVGLVLMODLAFISGFPPTFKET 240
QY 241 NQKTKLPENLSSKVKLLQLYSEASVALKLNPKDFOELNKQTKKMTIDGKELTISPA 300
DB 241 NQKTKLPENLSSKVKLLQLYSEASVALKLNPKDFOELNKQTKKMTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLRSIERHAPVWRNIPIYVNGQIPSWNL 360
DB 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLRSIERHAPVWRNIPIYVNGQIPSWNL 360
QY 361 DNPRTVITVHQDVFNRLSHLPTSSPAIESHIRIEGLSQKFTYLNDDVMFGKDWPPDF 420
DB 361 DNPRTVITVHQDVFNRLSHLPTSSPAIESHIRIEGLSQKFTYLNDDVMFGKDWPPDF 420
QY 421 YSHSGQKQYVLTWVPVNCAGCGPGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYTAG 480
DB 421 YSHSGQKQYVLTWVPVNCAGCGPGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYTAG 480
QY 481 GGGTGSIGVGHPMQFGGGINSVYCNQGCANSMADKFCDOACNVLSGCFDAGDCQDHF 540
DB 481 GGGTGSIGVGHPMQFGGGINSVYCNQGCANSMADKFCDOACNVLSGCFDAGDCQDHF 540
QY 541 HELYKYLIPNQTHYIIPKGECLPYPSFAEVAKRGVGAYSNDPIIRHASIANKKTIHL 600
```


Db 541 HELYKYLIPNQHYYIIPKGECLPYPSFAEVAKGVEGAYSDNPIIRHASIANKKTIHL 600
QY 601 IMHSGNATTIHENLTFQNTNDEEFKMOITVEVDTRREGKLNSTAKGYENLVSPITLLP 660
Db 601 IMHSGNATTIHENLTFQNTNDEEFKMOITVEVDTRREGKLNSTAKGYENLVSPITLLP 660
QY 661 EAEILFEDIPEKGRPFKFRHDVNSTRAQOEVKIPLVNI SLPPDAQLSANTLDLOLEH 720
Db 661 EAEILFEDIPEKGRPFKFRHDVNSTRAQOEVKIPLVNI SLPPDAQLSANTLDLOLEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIITDEFNDSLVAPQEKQVHKSIIPNSIG 780
Db 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIITDEFNDSLVAPQEKQVHKSIIPNSIG 780
QY 781 VSEIRLQRLTFPAPVAVKVNHDGQNPDLLETTAFRVETHQKTIIGNTKKEKPPSLIV 840
Db 781 VSEIRLQRLTFPAPVAVKVNHDGQNPDLLETTAFRVETHQKTIIGNTKKEKPPSLIV 840
QY 841 PLESQWTKKKTITGKEKENSMEENAEHNIIGTEVLGRKLOHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQWTKKKTITGKEKENSMEENAEHNIIGTEVLGRKLOHYTDSYLGFLPWEKKKYF 900
QY 901 QDLDEESLKTQLAFTDSKNTGRQK 928
Db 901 QDLDEESLKTQLAFTDSKNTGRQK 928
RESULT 5
US-10-023-890-4
; Sequence 4, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYD
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-890-4
Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MFKLQROTTCCLSHRYGLVYCFGVVVTIVSAFOGEBVLEMSRDQYHVLFDYSYRNI 60
Db 1 MFKLQROTTCCLSHRYGLVYCFGVVVTIVSAFOGEBVLEMSRDQYHVLFDYSYRNI 60
QY 61 AGKSPQNRCLPMPIDVVYTWNGTDLLELKELOQVREOMEBOQXAMREILKNTTEPTK 120
Db 61 AGKSPQNRCLPMPIDVVYTWNGTDLLELKELOQVREOMEBOQXAMREILKNTTEPTK 120
QY 121 KSEKQLECLLTHCIRKPMVLVDPALPANITLKDVPSLPSRHSASDINNAKPKPSINV 180
Db 121 KSEKQLECLLTHCIRKPMVLVDPALPANITLKDVPSLPSRHSASDINNAKPKPSINV 180
QY 181 SYVPEPSTKDVEDASGLKNGSRQTVWNGYLTDTKEVGLVIMODLAFLSGFPPTFKET 240
Db 181 SYVPEPSTKDVEDASGLKNGSRQTVWNGYLTDTKEVGLVIMODLAFLSGFPPTFKET 240
QY 241 NOLKTKLPENLSSKRYKLLQYSEASVALIKLNNPQFQELNKQYKNNITDGBLITSPA 300
Db 241 NOLKTKLPENLSSKRYKLLQYSEASVALIKLNNPQFQELNKQYKNNITDGBLITSPA 300
QY 301 YLLMPLSAISQKODEDISASRFEDNEELRYSLSISIERHAPVVRNIFIVTNGQIIPSWML 360
Db 301 YLLMPLSAISQKODEDISASRFEDNEELRYSLSISIERHAPVVRNIFIVTNGQIIPSWML 360

QY 361 DNPRTVITHQDVERNLSHLPTSSPAIESHTRIEGSSQKFIYLNDDVMPFGKDVPPDF 420
Db 361 DNPRTVITHQDVERNLSHLPTSSPAIESHTRIEGSSQKFIYLNDDVMPFGKDVPPDF 420
QY 421 YSHSKQKXYLLTWPEVNCAGCGPGSWIKDGYCDKACNNASACMDGDCGSGNSGRYIAG 480
Db 421 YSHSKQKXYLLTWPEVNCAGCGPGSWIKDGYCDKACNNASACMDGDCGSGNSGRYIAG 480
QY 481 GGGTGSITGCHPWQFQGGINSVYCNQGCANWLDKFCQACNVLSGCFDAGDGGQDHF 540
Db 481 GGGTGSITGCHPWQFQGGINSVYCNQGCANWLDKFCQACNVLSGCFDAGDGGQDHF 540
QY 541 HELYKYLIPNQHYYIIPKGECLPYPSFAEVAKGVEGAYSDNPIIRHASIANKKTIHL 600
Db 541 HELYKYLIPNQHYYIIPKGECLPYPSFAEVAKGVEGAYSDNPIIRHASIANKKTIHL 600
QY 601 IMHSGNATTIHENLTFQNTNDEEFKMOITVEVDTRREGKLNSTAKGYENLVSPITLLP 660
Db 601 IMHSGNATTIHENLTFQNTNDEEFKMOITVEVDTRREGKLNSTAKGYENLVSPITLLP 660
QY 661 EAEILFEDIPEKGRPFKFRHDVNSTRAQOEVKIPLVNI SLPPDAQLSANTLDLOLEH 720
Db 661 EAEILFEDIPEKGRPFKFRHDVNSTRAQOEVKIPLVNI SLPPDAQLSANTLDLOLEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIITDEFNDSLVAPQEKQVHKSIIPNSIG 780
Db 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIITDEFNDSLVAPQEKQVHKSIIPNSIG 780
QY 781 VSEIRLQRLTFPAPVAVKVNHDGQNPDLLETTAFRVETHQKTIIGNTKKEKPPSLIV 840
Db 781 VSEIRLQRLTFPAPVAVKVNHDGQNPDLLETTAFRVETHQKTIIGNTKKEKPPSLIV 840
QY 841 PLESQWTKKKTITGKEKENSMEENAEHNIIGTEVLGRKLOHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQWTKKKTITGKEKENSMEENAEHNIIGTEVLGRKLOHYTDSYLGFLPWEKKKYF 900
QY 901 QDLDEESLKTQLAFTDSKNTGRQK 928
Db 901 QDLDEESLKTQLAFTDSKNTGRQK 928
RESULT 6
US-10-024-197-4
; Sequence 4, Application US/10024197
; Publication No. US20030133924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREOSIDASE AND METHODS
; FILE REFERENCE: 209794US0
; CURRENT APPLICATION NUMBER: US/10/024,197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-024-197-4
Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MFKLQROTTCCLSHRYGLVYCFGVVVTIVSAFOGEBVLEMSRDQYHVLFDYSYRNI 60
Db 1 MFKLQROTTCCLSHRYGLVYCFGVVVTIVSAFOGEBVLEMSRDQYHVLFDYSYRNI 60
QY 61 AGKSPQNRCLPMPIDVVYTWNGTDLLELKELOQVREOMEBOQXAMREILKNTTEPTK 120
Db 61 AGKSPQNRCLPMPIDVVYTWNGTDLLELKELOQVREOMEBOQXAMREILKNTTEPTK 120

QY	121	USEKQTECLLHCICVPEVLVLDPALPATITLKQVPSLXPSHSHASDINFWAKPKPKPSINV	180
Dp	121	KSEKQTECLLHCICVPEVLVLDPALPATITLKQVPSLXPSHSHASDINFWAKPKPKPSINV	180
QY	181	SVVVPFDSTKQVEDAHSGLLKGNRSQTVWRGVLTTDKVEYGVJLMODLAFLSGFPEPTKET	240
Dp	181	SVVVPFDSTKQVEDAHSGLLKGNRSQTVWRGVLTTDKVEYGVJLMODLAFLSGFPEPTKET	240
QY	241	NQKTKLPEBNTSSKYKLLQLYSEASVALLKLNPKDFOELNKQTKQNTIDGKEITISPA	300
Dp	241	NQKTKLPEBNTSSKYKLLQLYSEASVALLKLNPKDFOELNKQTKQNTIDGKEITISPA	300
QY	301	YLMDLSAISQSKODEDISASRFEDEBELRSLSRSEHAWVRNIFVTNGQIPSWNTL	360
Dp	301	YLMDLSAISQSKODEDISASRFEDEBELRSLSRSEHAWVRNIFVTNGQIPSWNTL	360
QY	361	DNPRVTIYTHQDVERNLSHLPTFSSPAIESHHRIEGLSOKFIYLANDVMEGKWPPDF	420
Dp	361	DNPRVTIYTHQDVERNLSHLPTFSSPAIESHHRIEGLSOKFIYLANDVMEGKWPPDF	420
QY	421	YSHSKGQVNYLTWPRPNCAGCCPGSMIDGTCYCDRACNNASACDMGDCGSGNSGSGRYTAG	480
Dp	421	YSHSKGQVNYLTWPRPNCAGCCPGSMIDGTCYCDRACNNASACDMGDCGSGNSGSGRYTAG	480
QY	481	GGGGSISVGHPMQGGGINSVSYCNOGCANSWLADKFCDOACNVLSGCPGAGDGGDHF	540
Dp	481	GGGGSISVGHPMQGGGINSVSYCNOGCANSWLADKFCDOACNVLSGCPGAGDGGDHF	540
QY	541	HELYKVVLLPNQTHYIIPKGCCLPYFSPAEVAKRGVEGAYSNDPIIRHASIANKKTIHL	600
Dp	541	HELYKVVLLPNQTHYIIPKGCCLPYFSPAEVAKRGVEGAYSNDPIIRHASIANKKTIHL	600
QY	601	IMHSGMNAATTIHFNLTFOQNTDEEFKQIITVEVDREBPKINSTQKYEMLVSPITLLP	660
Dp	601	IMHSGMNAATTIHFNLTFOQNTDEEFKQIITVEVDREBPKINSTQKYEMLVSPITLLP	660
QY	661	EAEILLFEDIPKEKRFPKFKRHDVNSTRAQEBVKIPLVNISLLPKDQALSTNTLDLQEH	720
Dp	661	EAEILLFEDIPKEKRFPKFKRHDVNSTRAQEBVKIPLVNISLLPKDQALSTNTLDLQEH	720
QY	721	GDITLKGYNLSKSLALNSFLMNSQHAKIKNOAITTDETNDSLVAPOEKQVHKSILPNSLG	780
Dp	721	GDITLKGYNLSKSLALNSFLMNSQHAKIKNOAITTDETNDSLVAPOEKQVHKSILPNSLG	780
QY	781	VSESRJORTPBAVSAKXVNGHDQONPDLDTTARFRETHTOKTIGGNTKEXPPSLIY	840
Dp	781	VSESRJORTPBAVSAKXVNGHDQONPDLDTTARFRETHTOKTIGGNTKEXPPSLIY	840
QY	841	PLBSQMTKEKKITGKEKENSMEBEAENHIVTEVTLGRKLOHTDYSYLAGLPEPKKKYF	900
Dp	841	PLBSQMTKEKKITGKEKENSMEBEAENHIVTEVTLGRKLOHTDYSYLAGLPEPKKKYF	900
QY	901	QDLDDEBSLKTOLAYFTDSKNTGRQK 928	
Dp	901	QDLDDEBSLKTOLAYFTDSKNTGRQK 928	

	TYPE: PRT	ORGANISM: Homo sapiens	Query Match	100.0%; Score 4907; DB 14; Length 928;	Best Local Similarity 100.0%; Pred. No. 0;	Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY	1	MLFKLQRTYTCLSHRGYLYVCLGVVVTIVSAFOFGEVYLBNRSDQYHVLFDSDYDNI	60			
Db	1	MLFKLQRTYTCLSHRGYLYVCLGVVVTIVSAFOFGEVYLBNRSDQYHVLFDSDYDNI	60			
QY	61	AGSKQNNLCIMPLIDVYTVVNGDLELLKELOVRQMEEEQCAMBEILLGKNTTEPRK	120			
Db	61	AGSKQNNLCIMPLIDVYTVVNGDLELLKELOVRQMEEEQCAMBEILLGKNTTEPRK	120			
QY	121	KSEKQECTLTHCIVKPMVLVDPALPANITLKDVSLPSPSHASDINVAKPKNPSTNV	180			
Db	121	KSEKQECTLTHCIVKPMVLVDPALPANITLKDVSLPSPSHASDINVAKPKNPSTNV	180			
QY	181	SVVVPDSTKVDADASHGLKGNKQTVWRGTLTDDKVEYVGLVMDLAFLSGPPTEKET	240			
Db	181	SVVVPDSTKVDADASHGLKGNKQTVWRGTLTDDKVEYVGLVMDLAFLSGPPTEKET	240			
QY	241	NOLTKLPENLSSKYKLLQLYSEASVALKLKNPDPQELNKKQTKGNNTIGKEKLTISPA	300			
Db	241	NOLTKLPENLSSKYKLLQLYSEASVALKLKNPDPQELNKKQTKGNNTIGKEKLTISPA	300			
QY	301	YLLMDLSAISGSKODEDISASRFEDNEELRYSLRSIERHAPVWNIPIVINGQIIPSWNL	360			
Db	301	YLLMDLSAISGSKODEDISASRFEDNEELRYSLRSIERHAPVWNIPIVINGQIIPSWNL	360			
QY	361	DNPRATVTHQDVPRNLSHLTFESSPAIESHIHREBGSQFETYNDDVMGKVWPDPF	420			
Db	361	DNPRATVTHQDVPRNLSHLTFESSPAIESHIHREBGSQFETYNDDVMGKVWPDPF	420			
QY	421	YSHSKQGVYLTWPVYPCABEGCPGSMIDKGCDRACNNASACDMDGDCSGNSGGSRYTAG	480			
Db	421	YSHSKQGVYLTWPVYPCABEGCPGSMIDKGCDRACNNASACDMDGDCSGNSGGSRYTAG	480			
QY	481	GGGRTSIVGHPNPGGGINSVSYTCNCGCANSMLADKCDQACNVLSGCPDAGDCGDHF	540			
Db	481	GGGRTSIVGHPNPGGGINSVSYTCNCGCANSMLADKCDQACNVLSGCPDAGDCGDHF	540			
QY	541	HELYKVLILPQUTHIIPKGECLYFSPFAVAKGVEBAYSQNPITIRASJANKKTIHL	600			
Db	541	HELYKVLILPQUTHIIPKGECLYFSPFAVAKGVEBAYSQNPITIRASJANKKTIHL	600			
QY	601	IMHSGMNAATTIHNLTFQNTNDEEPRKQITVEVDTRBGPKLNSTAQKYNLVSPITLLP	660			
Db	601	IMHSGMNAATTIHNLTFQNTNDEEPRKQITVEVDTRBGPKLNSTAQKYNLVSPITLLP	660			
QY	661	EAILTFEDIPKEXKPPKFKRDVNSTRAQGEVXIPLVNISSLPRDQSLNTLDDLOEH	720			
Db	661	EAILTFEDIPKEXKPPKFKRDVNSTRAQGEVXIPLVNISSLPRDQSLNTLDDLOEH	720			
QY	721	GDITLKGYNLSKSLNSFLMNSQHAKIXQAIITDETNDSLVAPQEKQVHKSILPNSLG	780			
Db	721	GDITLKGYNLSKSLNSFLMNSQHAKIXQAIITDETNDSLVAPQEKQVHKSILPNSLG	780			
QY	781	VSERLQRTTFPAVSVKNGHQQGNPPDLDTTARFRVETHTQKTIGANTKEKPPSLIV	840			
Db	781	VSERLQRTTFPAVSVKNGHQQGNPPDLDTTARFRVETHTQKTIGANTKEKPPSLIV	840			
QY	841	PLESQMTKEKKITTEKEXKNSMEENAENHIGVTEVLGKRKLQHTDYSYLGLPMEKKXYF	900			
Db	841	PLESQMTKEKKITTEKEXKNSMEENAENHIGVTEVLGKRKLQHTDYSYLGLPMEKKXYF	900			
QY	901	QDLIDDEBSLKTQALAYFTDSKNTGRQK 928				
Db	901	QDLIDDEBSLKTQALAYFTDSKNTGRQK 928				

```

RESULT 8
US-10-306-686-1
; Sequence 1, Application US/10306686
; Publication No. US20030148460A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANase OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 230397US77DIV
; CURRENT APPLICATION NUMBER: US/10/306,686
; CURRENT FILING DATE: 2002-11-29
; PRIOR APPLICATION NUMBER: 09/636,596
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-08-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-306-686-1

```

```

Query Match      100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 MFPLLQRTYTCLSHRIGLYCPIGVVVTIYSAFQFGSVLEMSRDQYHVLFDPSYRDN 60
DB      1 MFPLLQRTYTCLSHRIGLYCPIGVVVTIYSAFQFGSVLEMSRDQYHVLFDPSYRDN 60
QY      61 AGKSFQNRCLPMPIDVYVTVNGTDLLELKELOQVREOMEBOQAKSEIIGKNTTPTK 120
DB      61 AGKSFQNRCLPMPIDVYVTVNGTDLLELKELOQVREOMEBOQAKSEIIGKNTTPTK 120
QY      121 KSEKOLECLTHCICIVPMVLDPALPANITLKDVPSELVPSFASDIFNVAKPKKPNSTNV 180
DB      121 KSEKOLECLTHCICIVPMVLDPALPANITLKDVPSELVPSFASDIFNVAKPKKPNSTNV 180
QY      121 KSEKOLECLTHCICIVPMVLDPALPANITLKDVPSELVPSFASDIFNVAKPKKPNSTNV 180
DB      121 KSEKOLECLTHCICIVPMVLDPALPANITLKDVPSELVPSFASDIFNVAKPKKPNSTNV 180
QY      181 SVVVPDSTKDVDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLMODLAFSGPPTPET 240
DB      181 SVVVPDSTKDVDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLMODLAFSGPPTPET 240
QY      241 NOLKTKLPENLSSKYLLOLYSEASVALLKLNPKDFOLNKQTKKNTTIDGKELTISPA 300
DB      241 NOLKTKLPENLSSKYLLOLYSEASVALLKLNPKDFOLNKQTKKNTTIDGKELTISPA 300
QY      301 YLLMDLSAISQKQEDISASRFEDNEELRYSLSIRERHAPWVRNIFVTNGQIPSWLNL 360
DB      301 YLLMDLSAISQKQEDISASRFEDNEELRYSLSIRERHAPWVRNIFVTNGQIPSWLNL 360
QY      361 DNPRTVITVHOVFNRLSHLPFSSPAIESHRIEGLSOKFIYLLNDVDFGKDVMPDDF 420
DB      361 DNPRTVITVHOVFNRLSHLPFSSPAIESHRIEGLSOKFIYLLNDVDFGKDVMPDDF 420
QY      421 YHSHSQQKXYLLTPVPCNACGCGPSWITDGYCDKACNNSACDMGDCGSGSGSRIYAG 480
DB      421 YHSHSQQKXYLLTPVPCNACGCGPSWITDGYCDKACNNSACDMGDCGSGSGSRIYAG 480
QY      481 GGTGSSIGVGHPMQGGGINSYCNQGCANSMWLDKFCDOACNVLSGCFPAGCGODHF 540
DB      481 GGTGSSIGVGHPMQGGGINSYCNQGCANSMWLDKFCDOACNVLSGCFPAGCGODHF 540
QY      541 HELYKYLIPNQHYYIIPKGECLPYPSFAVAKRGVEGAYSDNPIIRHSAIANMKKTIHL 600
DB      541 HELYKYLIPNQHYYIIPKGECLPYPSFAVAKRGVEGAYSDNPIIRHSAIANMKKTIHL 600
QY      601 INHSGMNAATIHENLTFOINTDEEFMOTIYEVDPREGPKLNSTOKXEYLVSPITLLP 660
DB      601 INHSGMNAATIHENLTFOINTDEEFMOTIYEVDPREGPKLNSTOKXEYLVSPITLLP 660
QY      661 EAEIIFEDIPEKEKPPKRGHDVNSTRAOEYVKIPLVNIISLPEKDAQLSINTLDLOEH 720
DB      661 EAEIIFEDIPEKEKPPKRGHDVNSTRAOEYVKIPLVNIISLPEKDAQLSINTLDLOEH 720

```

```

QY      721 GDITLKGYNLSKSGALLRSFLMNSQHAKIKNOAITDETNDLSVAPQEKOVHKSIIIPNSIG 780
DB      721 GDITLKGYNLSKSGALLRSFLMNSQHAKIKNOAITDETNDLSVAPQEKOVHKSIIIPNSIG 780
QY      781 VSRRLQRLFPFPAVSVKNGHDOQNPDLLETTAFRVVTHQKTIIGVNTKEKPPSLIV 840
DB      781 VSRRLQRLFPFPAVSVKNGHDOQNPDLLETTAFRVVTHQKTIIGVNTKEKPPSLIV 840
QY      841 PLESQMTKEKKTIGKEKENSMEENAEHNIIGVTEVLLGRKLOHYTDSYLGFLPWEKKYF 900
DB      841 PLESQMTKEKKTIGKEKENSMEENAEHNIIGVTEVLLGRKLOHYTDSYLGFLPWEKKYF 900
QY      901 QDLLEDEESIKTQLAYFTDSKNTGRQK 928
DB      901 QDLLEDEESIKTQLAYFTDSKNTGRQK 928

```

```

RESULT 9
US-10-023-888-2
; Sequence 2, Application US/10023888
; Publication No. US20030119088A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: SOLUBLE GLUCAN PHOSPHOTRANSFERASE
; FILE REFERENCE: 203515US77
; CURRENT APPLICATION NUMBER: US/10/023,888
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: hybrid
US-10-023-888-2

```

```

Query Match      94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      45 SRDQHVLPDSYRDNIAAGSFQNRCLPMPIDVYVTVNGTDLLELKELOQVREOMEBOQ 104
DB      35 SRDQHVLPDSYRDNIAAGSFQNRCLPMPIDVYVTVNGTDLLELKELOQVREOMEBOQ 94
QY      105 KAMREILGKNTTEPTKSEKOLECLTHCICIVPMVLDPALPANITLKDVPSELVPSFSA 164
DB      95 KAMREILGKNTTEPTKSEKOLECLTHCICIVPMVLDPALPANITLKDVPSELVPSFSA 154
QY      165 SDIFNVAKPKPNSTNVSVVVPDSTKDVDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLM 224
DB      155 SDIFNVAKPKPNSTNVSVVVPDSTKDVDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLM 214
QY      225 QDLAFSGPPTPETKENTNOLKTKLPENLSSKYLLOLYSEASVALLKLNPKDFOLNKQ 284
DB      215 QDLAFSGPPTPETKENTNOLKTKLPENLSSKYLLOLYSEASVALLKLNPKDFOLNKQ 274
QY      285 KKNMTIDGKELTISPAYLLMDLSAISQKQEDISASRFEDNEELRYSLSIRERHAPWVR 344
DB      275 KKNMTIDGKELTISPAYLLMDLSAISQKQEDISASRFEDNEELRYSLSIRERHAPWVR 334
QY      345 NTFIVNGQIPSWLNDNPRVITVHOVFNRLSHLPFSSPAIESHRIEGLSOKFIY 404
DB      335 NTFIVNGQIPSWLNDNPRVITVHOVFNRLSHLPFSSPAIESHRIEGLSOKFIY 394
QY      405 LNDVAFGADVMPDDFYSHSGQKXYLLTPVPCNACGCGPSWITDGYCDKACNNSACDM 464
DB      395 LNDVAFGADVMPDDFYSHSGQKXYLLTPVPCNACGCGPSWITDGYCDKACNNSACDM 454
QY      465 GGDGSGSGSGSRIYAGGGTGSIGVGHPMQGGGINSYCNQGCANSMWLDKFCDOACN 524
DB      455 GGDGSGSGSGSRIYAGGGTGSIGVGHPMQGGGINSYCNQGCANSMWLDKFCDOACN 514
QY      525 VLSGCFDADCGODHFHELXYKYLIPNQHYYIIPKGECLPYPSFAVAKRGVEGAYSDNP 584

```

```
Db 515 VLSGCGDADCCGDHHELYKVIILPNQTHYIIPKGECLPYSFSAEVAKRGEAGSDNP 574
Qy 585 IIRHASIANKMTIHLIMSGNATTIHFNLTFOKTNDEBFKQITVEVDTRREGPKLNT 644
Db 575 IIRHASIANKMTIHLIMSGNATTIHFNLTFOKTNDEBFKQITVEVDTRREGPKLNT 634
Qy 645 AOKGYENLVSPITLLPEAEILFEDIPEKRFKFRKHDVNSTRAQSEVKIPLVNIISLLP 704
Db 635 AOKGYENLVSPITLLPEAEILFEDIPEKRFKFRKHDVNSTRAQSEVKIPLVNIISLLP 694
Qy 705 KDAQSLNTLDLQLEHGDITLKGYNLKSALLRSLFMSQAKIKNOAIIITDETNDSLVA 764
Db 695 KDAQSLNTLDLQLEHGDITLKGYNLKSALLRSLFMSQAKIKNOAIIITDETNDSLVA 754
Qy 765 POEKVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPDLLETTARFVEETHQK 824
Db 755 POEKVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPDLLETTARFVEETHQK 814
Qy 825 TIGGVNTEKRPSPILVPLESQMTKEKKTITGEKENSMBENAMNHIQVTEVLGRKLOHY 884
Db 815 TIGGVNTEKRPSPILVPLESQMTKEKKTITGEKENSMBENAMNHIQVTEVLGRKLOHY 874
Qy 885 TDSYLGFLPWEKKKTYFODLLDEESLKTOLAYFTDSKNTGR 925
Db 875 TDSYLGFLPWEKKKTYFODLLDEESLKTOLAYFTDSKNTGR 915
```

```
RESULT 10
US-10-023-889-2
; Sequence 2, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBOH
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRF
; ORGANISM: hybrid
US-10-023-889-2
```

```
Query Match 94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 45 SRDQYHVLFDPSYRDNIAKSFQNRCLPMPIDVYVTWNGTDLELLKELQVREQMEEQ 104
Db 35 SRDQYHVLFDPSYRDNIAKSFQNRCLPMPIDVYVTWNGTDLELLKELQVREQMEEQ 94
Qy 105 KAMREILGNKTEPTTKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISYPSFSA 164
Db 95 KAMREILGNKTEPTTKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISYPSFSA 154
Qy 165 SDIFNVAKPKNSTNVSVVVFSTKDVEDAHSGLLKNSRQVWARGYITTDKEVPGVLVM 224
Db 155 SDIFNVAKPKNSTNVSVVVFSTKDVEDAHSGLLKNSRQVWARGYITTDKEVPGVLVM 214
Qy 225 QDLAFISGPPFPEKTEKQLEKTLKPLPENLSKXVLLQLYSEASVALLKLNPKDFOLNKT 284
Db 215 QDLAFISGPPFPEKTEKQLEKTLKPLPENLSKXVLLQLYSEASVALLKLNPKDFOLNKT 274
Qy 285 KKNMTIDGKELTISPAYLLMDLSAISQKQDDDISASRPEDNEELRYSLSIERHAPVVR 344
Db 275 KKNMTIDGKELTISPAYLLMDLSAISQKQDDDISASRPEDNEELRYSLSIERHAPVVR 334
Qy 345 NFIIVTNGQIPSMNLNDNPRVTIIVTHQDVFRLNLSHPFSSPAIESHRIEGLSQKFIY 404
Db 335 NFIIVTNGQIPSMNLNDNPRVTIIVTHQDVFRLNLSHPFSSPAIESHRIEGLSQKFIY 394
```

```
Db 335 NFIIVTNGQIPSMNLNDNPRVTIIVTHQDVFRLNLSHPFSSPAIESHRIEGLSQKFIY 394
Qy 405 LNDVWVGKDVMPDDPFYSHSKGQKVVLTWVPNCAGECPSWMTKDQYCDKACNNASCDMD 464
Db 395 LNDVWVGKDVMPDDPFYSHSKGQKVVLTWVPNCAGECPSWMTKDQYCDKACNNASCDMD 454
Qy 465 GSDCSGNSGGSRYYIAGGGTGSIGVHPWQFGGINSVSVCNCGCANSWLADFCQACN 524
Db 455 GSDCSGNSGGSRYYIAGGGTGSIGVHPWQFGGINSVSVCNCGCANSWLADFCQACN 514
Qy 525 VLSGCPDAGCGGDHHELYKVIILPNQTHYIIPKGECLPYSFSAEVAKRGEAGASDNP 584
Db 515 VLSGCPDAGCGGDHHELYKVIILPNQTHYIIPKGECLPYSFSAEVAKRGEAGASDNP 574
Qy 585 IIRHASIANKMTIHLIMSGNATTIHFNLTFOKTNDEBFKQITVEVDTRREGPKLNT 644
Db 575 IIRHASIANKMTIHLIMSGNATTIHFNLTFOKTNDEBFKQITVEVDTRREGPKLNT 634
Qy 645 AOKGYENLVSPITLLPEAEILFEDIPEKRFKFRKHDVNSTRAQSEVKIPLVNIISLLP 704
Db 635 AOKGYENLVSPITLLPEAEILFEDIPEKRFKFRKHDVNSTRAQSEVKIPLVNIISLLP 694
Qy 705 KDAQSLNTLDLQLEHGDITLKGYNLKSALLRSLFMSQAKIKNOAIIITDETNDSLVA 764
Db 695 KDAQSLNTLDLQLEHGDITLKGYNLKSALLRSLFMSQAKIKNOAIIITDETNDSLVA 754
Qy 765 POEKVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPDLLETTARFVEETHQK 824
Db 755 POEKVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPDLLETTARFVEETHQK 814
Qy 825 TIGGVNTEKRPSPILVPLESQMTKEKKTITGEKENSMBENAMNHIQVTEVLGRKLOHY 884
Db 815 TIGGVNTEKRPSPILVPLESQMTKEKKTITGEKENSMBENAMNHIQVTEVLGRKLOHY 874
Qy 885 TDSYLGFLPWEKKKTYFODLLDEESLKTOLAYFTDSKNTGR 925
Db 875 TDSYLGFLPWEKKKTYFODLLDEESLKTOLAYFTDSKNTGR 915
```

```
RESULT 11
US-10-023-890-2
; Sequence 2, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOH
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRF
; ORGANISM: hybrid
US-10-023-890-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 45 SRDQYHVLFDPSYRDNIAKSFQNRCLPMPIDVYVTWNGTDLELLKELQVREQMEEQ 104
Db 35 SRDQYHVLFDPSYRDNIAKSFQNRCLPMPIDVYVTWNGTDLELLKELQVREQMEEQ 94
Qy 105 KAMREILGNKTEPTTKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISYPSFSA 164
Db 95 KAMREILGNKTEPTTKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISYPSFSA 154
Qy 165 SDIFNVAKPKNSTNVSVVVFSTKDVEDAHSGLLKNSRQVWARGYITTDKEVPGVLVM 224
Db 155 SDIFNVAKPKNSTNVSVVVFSTKDVEDAHSGLLKNSRQVWARGYITTDKEVPGVLVM 214
```

QY 225 QDLAFLSGPPPTKEKNOLKTKL PENLSKVKLLQLYSASVALLKNNPKDFOELNKOT 284
 DB 215 QDLAFLSGPPPTKEKNOLKTKL PENLSKVKLLQLYSASVALLKNNPKDFOELNKOT 274
 QY 285 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEBELRSLRSIERHAPVR 344
 DB 275 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEBELRSLRSIERHAPVR 334
 QY 345 NFIYVNGQIPSWMLNDNPRVTI VTHQVFNLSHLPTSSPAIESHTRIEGLSOKFIY 404
 DB 335 NFIYVNGQIPSWMLNDNPRVTI VTHQVFNLSHLPTSSPAIESHTRIEGLSOKFIY 394
 QY 405 LNDVWFGKDWPDDEFYSHSKQKXYLLTPVNCABGCGSWIKDGYCDKACNNSACMD 464
 DB 395 LNDVWFGKDWPDDEFYSHSKQKXYLLTPVNCABGCGSWIKDGYCDKACNNSACMD 454
 QY 465 GDDCGNSGSGRYIAGGGGTGSI GVGHPMQFGGINSVSYCNOGANSMLADKFCDOACN 524
 DB 455 GDDCGNSGSGRYIAGGGGTGSI GVGHPMQFGGINSVSYCNOGANSMLADKFCDOACN 514
 QY 525 VLSGCFDAGDCQGDHFEHLKYVILLPNQTHYIIPGECCLPYSPFAVAKRGVAGAYSDNP 584
 DB 515 VLSGCFDAGDCQGDHFEHLKYVILLPNQTHYIIPGECCLPYSPFAVAKRGVAGAYSDNP 574
 QY 585 IIRHASIAKMKTIHLIMSGMNAATTIHNLTFONTNDEEFKQITVEVDTRREGKLNST 644
 DB 575 IIRHASIAKMKTIHLIMSGMNAATTIHNLTFONTNDEEFKQITVEVDTRREGKLNST 634
 QY 645 AOKGYENLVSPITLLPEAEILFEDI PKERKPFKFRHDVNSRRAROEVEKILPVNISLLP 704
 DB 635 AOKGYENLVSPITLLPEAEILFEDI PKERKPFKFRHDVNSRRAROEVEKILPVNISLLP 694
 QY 705 KQAQSLNTLDDQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA 764
 DB 695 KQAQSLNTLDDQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA 754
 QY 765 POEKQYHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPPLDLETTARFRVETHQX 824
 DB 755 POEKQYHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPPLDLETTARFRVETHQX 814
 QY 825 TIGGANTYKKEPSPSLIVPLESOMTKKTKTGKKEKNSRMEENHIGTEVILGKRLQHY 884
 DB 815 TIGGANTYKKEPSPSLIVPLESOMTKKTKTGKKEKNSRMEENHIGTEVILGKRLQHY 874
 QY 885 TDSYLGFLPWEKKKXYFODLLDEESLKTOLAYFTDSKNTGR 925
 DB 875 TDSYLGFLPWEKKKXYFODLLDEESLKTOLAYFTDSKNTGR 915

RESULT 12
 US-10-024-197-2
 ; Sequence 2, Application US/10024197
 ; Publication No. US20030133924A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, William
 ; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS C
 ; FILE REFERENCE: 209794USO
 ; CURRENT APPLICATION NUMBER: US/10/024,197
 ; CURRENT FILING DATE: 2001-12-21
 ; NUMBER OF SEQ ID NOS: 27
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 1199
 ; TYPE: PRF
 ; ORGANISM: hybrid
 US-10-024-197-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;
 Best Local Similarity 99.3%; Preq. No. 0;
 Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SRDQYHVLFDPSYNDNIAKSFQNRCLCPMPIDVVTWVNGTDLBLIKELQOVRBEOMEEO 104
 DB 35 SRDQYHVLFDPSYNDNIAKSFQNRCLCPMPIDVVTWVNGTDLBLIKELQOVRBEOMEEO 94
 QY 105 KAMREILGKNTTEPTKSEKOLECLTHCICVPMULDALPANITLKXVPISLYSPFBSA 164
 DB 95 KAMREILGKNTTEPTKSEKOLECLTHCICVPMULDALPANITLKXVPISLYSPFBSA 154
 QY 165 SDIFNVAKPNPSTNSVVPFSTKQVEDASHGLKNGSRQVWAGYLLTDXEVPGLVIM 224
 DB 155 SDIFNVAKPNPSTNSVVPFSTKQVEDASHGLKNGSRQVWAGYLLTDXEVPGLVIM 214
 QY 225 QDLAFLSGPPPTKEKNOLKTKL PENLSKVKLLQLYSASVALLKNNPKDFOELNKOT 284
 DB 215 QDLAFLSGPPPTKEKNOLKTKL PENLSKVKLLQLYSASVALLKNNPKDFOELNKOT 274
 QY 285 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEBELRSLRSIERHAPVR 344
 DB 275 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEBELRSLRSIERHAPVR 334
 QY 345 NFIYVNGQIPSWMLNDNPRVTI VTHQVFNLSHLPTSSPAIESHTRIEGLSOKFIY 404
 DB 335 NFIYVNGQIPSWMLNDNPRVTI VTHQVFNLSHLPTSSPAIESHTRIEGLSOKFIY 394
 QY 405 LNDVWFGKDWPDDEFYSHSKQKXYLLTPVNCABGCGSWIKDGYCDKACNNSACMD 464
 DB 395 LNDVWFGKDWPDDEFYSHSKQKXYLLTPVNCABGCGSWIKDGYCDKACNNSACMD 454
 QY 465 GDDCGNSGSGRYIAGGGGTGSI GVGHPMQFGGINSVSYCNOGANSMLADKFCDOACN 524
 DB 455 GDDCGNSGSGRYIAGGGGTGSI GVGHPMQFGGINSVSYCNOGANSMLADKFCDOACN 514
 QY 525 VLSGCFDAGDCQGDHFEHLKYVILLPNQTHYIIPGECCLPYSPFAVAKRGVAGAYSDNP 584
 DB 515 VLSGCFDAGDCQGDHFEHLKYVILLPNQTHYIIPGECCLPYSPFAVAKRGVAGAYSDNP 574
 QY 585 IIRHASIAKMKTIHLIMSGMNAATTIHNLTFONTNDEEFKQITVEVDTRREGKLNST 644
 DB 575 IIRHASIAKMKTIHLIMSGMNAATTIHNLTFONTNDEEFKQITVEVDTRREGKLNST 634
 QY 645 AOKGYENLVSPITLLPEAEILFEDI PKERKPFKFRHDVNSRRAROEVEKILPVNISLLP 704
 DB 635 AOKGYENLVSPITLLPEAEILFEDI PKERKPFKFRHDVNSRRAROEVEKILPVNISLLP 694
 QY 705 KQAQSLNTLDDQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA 764
 DB 695 KQAQSLNTLDDQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA 754
 QY 765 POEKQYHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPPLDLETTARFRVETHQX 824
 DB 755 POEKQYHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPPLDLETTARFRVETHQX 814
 QY 825 TIGGANTYKKEPSPSLIVPLESOMTKKTKTGKKEKNSRMEENHIGTEVILGKRLQHY 884
 DB 815 TIGGANTYKKEPSPSLIVPLESOMTKKTKTGKKEKNSRMEENHIGTEVILGKRLQHY 874
 QY 885 TDSYLGFLPWEKKKXYFODLLDEESLKTOLAYFTDSKNTGR 925
 DB 875 TDSYLGFLPWEKKKXYFODLLDEESLKTOLAYFTDSKNTGR 915

RESULT 13
 US-10-023-894-2
 ; Sequence 2, Application US/10023894
 ; Publication No. US20030143669A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, William
 ; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
 ; FILE REFERENCE: 217139US77
 ; CURRENT APPLICATION NUMBER: US/10/023,894
 ; CURRENT FILING DATE: 2001-12-21

```
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: hybrid
US-10-023-894-2
```

```
Query Match      94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 45 SDDQYHVFDSYRDNIAGKSFONRLCLPMPIDVYTWNGDLELKELOQVREOMESEQ 104
DB 35 SDDQYHVFDSYRDNIAGKSFONRLCLPMPIDVYTWNGDLELKELOQVREOMESEQ 94
QY 105 KAMREILGKNTTEPTKSKSEKOLECLTHGCIKVPMLVLPALPANITLKDVPGLYSPFSA 164
DB 95 KAMREILGKNTTEPTKSKSEKOLECLTHGCIKVPMLVLPALPANITLKDVPGLYSPFSA 154
QY 165 SDIFVNAKXKNSSTNVSVVFPDSTKDVEDAHSGLLKXNSRQTVWRGYLTDDKEVPGVLM 224
DB 155 SDIFVNAKXKNSSTNVSVVFPDSTKDVEDAHSGLLKXNSRQTVWRGYLTDDKEVPGVLM 214
QY 225 ODLAFISGPPPTFKETNOLKTKLPENLSKXVLLQLYSEASVALLKANNPKDFOLNKOT 284
DB 215 ODLAFISGPPPTFKETNOLKTKLPENLSKXVLLQLYSEASVALLKANNPKDFOLNKOT 274
QY 285 KXNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYLSIRSIERRAPVVR 344
DB 275 KXNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYLSIRSIERRAPVVR 334
QY 345 NIFVTNGQIPSMNLNDNRVTIVTHQVFNLSHLPTFSSPAISHSIRIEGJSQKTIY 404
DB 335 NIFVTNGQIPSMNLNDNRVTIVTHQVFNLSHLPTFSSPAISHSIRIEGJSQKTIY 394
QY 405 LNDVVFAGDVPDDPFYSKQOKYLLTFVPVNCAGEGSGSIKXGYCDKACNSACMD 464
DB 395 LNDVVFAGDVPDDPFYSKQOKYLLTFVPVNCAGEGSGSIKXGYCDKACNSACMD 454
QY 465 GDDCGSNGSGSYIAGGGGTGSIQVGHPMQFGGINSVSYCQGCANSWLDKFCDOACN 524
DB 455 GDDCGSNGSGSYIAGGGGTGSIQVGHPMQFGGINSVSYCQGCANSWLDKFCDOACN 514
QY 525 VLSGCGPDAGDCGDHHEHLYKVTLLPNOGHTIIPKGECLPYSPFAVKRGVEGAYSDNP 584
DB 515 VLSGCGPDAGDCGDHHEHLYKVTLLPNOGHTIIPKGECLPYSPFAVKRGVEGAYSDNP 574
QY 585 IIRHASTANKMTIHLIMHSGNNAATTIHENLFTONTDEEPMQITVEVDTREBPKLNT 644
DB 575 IIRHASTANKMTIHLIMHSGNNAATTIHENLFTONTDEEPMQITVEVDTREBPKLNT 634
QY 645 AOKGYNVLSPTITLPEAEILFEDIPKERRPKFRHDVNSTRAQOEYKPLVNISSLP 704
DB 635 AOKGYNVLSPTITLPEAEILFEDIPKERRPKFRHDVNSTRAQOEYKPLVNISSLP 694
QY 705 KXAOULSINTLDELQLEHGDITLKGYNLSKSLRSEFAMNSQHAKINQOAIITDETDSLVA 764
DB 695 KXAOULSINTLDELQLEHGDITLKGYNLSKSLRSEFAMNSQHAKINQOAIITDETDSLVA 754
QY 765 POKROYHKSILPNSLGVSERLQRLTFPAVSXVNGHDQONPDLDETTARFVETHTOK 824
DB 755 POKROYHKSILPNSLGVSERLQRLTFPAVSXVNGHDQONPDLDETTARFVETHTOK 814
QY 825 TIGGANTKXKPSLIAPLESOMTKEKKITGKEKENSMBENAKENHIGTVEVILGKLOHY 884
DB 815 TIGGANTKXKPSLIAPLESOMTKEKKITGKEKENSMBENAKENHIGTVEVILGKLOHY 874
QY 885 TDSYLGFLPWEKKKYFODLLDEEESIKTOLAYFTDSKNTGR 925
DB 875 TDSYLGFLPWEKKKYFODLLDEEESIKTOLAYFTDSKNTGR 915
```

```
RESULT 14
US-10-120-801-88
; Sequence 88; Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Kerkuta, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Ruad
; APPLICANT: Topper, James N.
; APPLICANT: Malysanar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smitheon, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLES OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 88
; LENGTH: 847
; TYPE: PRT
; ORGANISM: human
US-10-120-801-88

Query Match      91.2%; Score 4473; DB 15; Length 847;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 845; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 MLEFLLORQTYTCLSHRYGLVCFGLGVVTVISAFQGEVLEWSRDQYHVFDSYRDNI 60
DB 1 MLEFLLORQTYTCLSHRYGLVCFGLGVVTVISAFQGEVLEWSRDQYHVFDSYRDNI 60
QY 61 AGKSFONRLCLPMPIDVYTWNGDLELKELOQVREOMESEQKAMREILGKNTTEPTK 120
DB 61 AGKSFONRLCLPMPIDVYTWNGDLELKELOQVREOMESEQKAMREILGKNTTEPTK 120
QY 121 KSEKOLECLTHGCIKVPMLVLPALPANITLKDVPGLYSPFHSASDIENYAKKNSSTNV 180
DB 121 KSEKOLECLTHGCIKVPMLVLPALPANITLKDVPGLYSPFHSASDIENYAKKNSSTNV 180
QY 181 SVAVFDSKDVEDAHSGLLKXNSRQTVWRGYLTDDKEVPLVMODLAFISGPPPTFKET 240
DB 181 SVAVFDSKDVEDAHSGLLKXNSRQTVWRGYLTDDKEVPLVMODLAFISGPPPTFKET 240
QY 241 NOLKTKLPENLSKXVLLQLYSEASVALLKANNPKDFOLNKOTKXNMTIDGKELTISPA 300
DB 241 NOLKTKLPENLSKXVLLQLYSEASVALLKANNPKDFOLNKOTKXNMTIDGKELTISPA 300
QY 301 YLMDLSAISQSKODEDISASRFEDNEELRYLSIRSIERRAPVVRNIFVTNGQIPSWLNI 360
DB 301 YLMDLSAISQSKODEDISASRFEDNEELRYLSIRSIERRAPVVRNIFVTNGQIPSWLNI 360
```


QY 361 DNRPRVITVHODVFRNLSHLPTFSSPAISHHRIEGLSOKETIYNDVDMFGKDWPDDE 420
DB 361 DNRPRVITVHODVFRNLSHLPTFSSPAISHHRIEGLSOKETIYNDVDMFGKDWPDDE 420
QY 421 YSHSKQKXYLLTWPVNCAGCGPGSWIKDGYCDKACNNACDWDGDCSGNSGSGRYTAG 480
DB 421 YSHSKQKXYLLTWPVNCAGCGPGSWIKDGYCDKACNNACDWDGDCSGNSGSGRYTAG 480
QY 481 GGGTSGIGVGHWPQFGGINSVSYCNQGCANSMWLDKFCDOACNVLSCGFDAGCGODHF 540
DB 481 GGGTSGIGVGHWPQFGGINSVSYCNQGCANSMWLDKFCDOACNVLSCGFDAGCGODHF 540
QY 541 HELYKYLIPNQTHYIIPKGECLPYPSFAEVAKRGEGAYSNDPIIRHAIANKKTIHL 600
DB 541 HELYKYLIPNQTHYIIPKGECLPYPSFAEVAKRGEGAYSNDPIIRHAIANKKTIHL 600
QY 601 IMHSGNNAATTIHFNLTFOQNTDEBFMOQITVEVDTRREGPKLNSTAKGEMLVSPITLLP 660
DB 601 IMHSGNNAATTIHFNLTFOQNTDEBFMOQITVEVDTRREGPKLNSTAKGEMLVSPITLLP 660
QY 661 EAEILFEDIPEKRRPFKFRHDVNSTRAQEEVKIPLVNIISLPPDOQLSNTLTLQLEH 720
DB 661 EAEILFEDIPEKRRPFKFRHDVNSTRAQEEVKIPLVNIISLPPDOQLSNTLTLQLEH 720
QY 721 GDTTLKGYNLSSKALRSFLMNSOHAQIKNOAIITDETNDLSVAPOEKOVHKSILPNSLG 780
DB 721 GDTTLKGYNLSSKALRSFLMNSOHAQIKNOAIITDETNDLSVAPOEKOVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKVNHDGQNPDLDETTAFRVETHOKTIGGNTVEKPSLIV 840
DB 781 VSERLQRLTFPAVSVKVNHDGQNPDLDETTAFRVETHOKTIGGNTVEKPSLIV 840
QY 841 PLESQMT 847
DB 841 PLESQMT 847

RESULT 15
US-10-094-749-2392
; Sequence 2392, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YUKI
; APPLICANT: OTSUKA, KAGURU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAMA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTOKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL, FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2392
; LENGTH: 846
; TYPE: PRT

; ORGANISM: Homo sapiens
US-10-094-749-2392
Query Match 91.1%; Score 4468; DB 15; Length 846;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 844; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MFRLQROQYTTGLSHRYGILYVCFLLGVVVTIYSAFQGVVLWMSRQYHVLFDSTRDNI 60
DB 1 MFRLQROQYTTGLSHRYGILYVCFLLGVVVTIYSAFQGVVLWMSRQYHVLFDSTRDNI 60
QY 61 AGSFQNRCLPPIPIVWVTVWNGTDLLELKELOQVREOMEEBOKMRRIIGKNTPEPK 120
DB 61 AGSFQNRCLPPIPIVWVTVWNGTDLLELKELOQVREOMEEBOKMRRIIGKNTPEPK 120
QY 121 KSEKQLECLTHCICIKPMVLDPALPANTLTKDLPSLYPSFHSASDIENVAKPKPSTNV 180
DB 121 KSEKQLECLTHCICIKPMVLDPALPANTLTKDLPSLYPSFHSASDIENVAKPKPSTNV 180
QY 181 SVVVPSTQDVEDNHSGLKGNRSQRTVMRGYLLTDKEVGVLMODLAFSGPPTFEKT 240
DB 181 SVVVPSTQDVEDNHSGLKGNRSQRTVMRGYLLTDKEVGVLMODLAFSGPPTFEKT 240
QY 241 NOLKTLPENLSKVKLLQLYSEASVALIKNNPKDFOELNQTKKMTIDGKELTISPA 300
DB 241 NOLKTLPENLSKVKLLQLYSEASVALIKNNPKDFOELNQTKKMTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKQDDEDSASREDNEELRYSLSIERAPVWRNIFTYNGQISMLNL 360
DB 301 YLLMDLSAISQSKQDDEDSASREDNEELRYSLSIERAPVWRNIFTYNGQISMLNL 360
QY 361 DNRPRVITVHODVFRNLSHLPTFSSPAISHHRIEGLSOKETIYNDVDMFGKDWPDDE 420
DB 361 DNRPRVITVHODVFRNLSHLPTFSSPAISHHRIEGLSOKETIYNDVDMFGKDWPDDE 420
QY 421 YSHSKQKXYLLTWPVNCAGCGPGSWIKDGYCDKACNNACDWDGDCSGNSGSGRYTAG 480
DB 421 YSHSKQKXYLLTWPVNCAGCGPGSWIKDGYCDKACNNACDWDGDCSGNSGSGRYTAG 480
QY 481 GGGTSGIGVGHWPQFGGINSVSYCNQGCANSMWLDKFCDOACNVLSCGFDAGCGODHF 540
DB 481 GGGTSGIGVGHWPQFGGINSVSYCNQGCANSMWLDKFCDOACNVLSCGFDAGCGODHF 540
QY 541 HELYKYLIPNQTHYIIPKGECLPYPSFAEVAKRGEGAYSNDPIIRHAIANKKTIHL 600
DB 541 HELYKYLIPNQTHYIIPKGECLPYPSFAEVAKRGEGAYSNDPIIRHAIANKKTIHL 600
QY 601 IMHSGNNAATTIHFNLTFOQNTDEBFMOQITVEVDTRREGPKLNSTAKGEMLVSPITLLP 660
DB 601 IMHSGNNAATTIHFNLTFOQNTDEBFMOQITVEVDTRREGPKLNSTAKGEMLVSPITLLP 660
QY 661 EAEILFEDIPEKRRPFKFRHDVNSTRAQEEVKIPLVNIISLPPDOQLSNTLTLQLEH 720
DB 661 EAEILFEDIPEKRRPFKFRHDVNSTRAQEEVKIPLVNIISLPPDOQLSNTLTLQLEH 720
QY 721 GDTTLKGYNLSSKALRSFLMNSOHAQIKNOAIITDETNDLSVAPOEKOVHKSILPNSLG 780
DB 721 GDTTLKGYNLSSKALRSFLMNSOHAQIKNOAIITDETNDLSVAPOEKOVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKVNHDGQNPDLDETTAFRVETHOKTIGGNTVEKPSLIV 840
DB 781 VSERLQRLTFPAVSVKVNHDGQNPDLDETTAFRVETHOKTIGGNTVEKPSLIV 840
QY 841 PLESQMT 846
DB 841 PLESQMT 846

Search completed: July 26, 2004, 11:25:25
Job time : 74.3389 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:58 ; Search time 15.4568 Seconds
(without alignments)

3126.212 Million cell updates/sec

Title: US-10-657-280-1

Perfect score: 4907
Sequence: 1 MFLKLRQRTYTCISHRYGL.....SLKTQLAYFTDSKNTGRQLK 928

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	168.5	3.4	2470	1 NTC2_MOUSE	O35516 mus musculus
2	168.5	3.4	2471	1 NTC2_HUMAN	Q04721 homo sapien
3	168.5	3.4	2471	1 NTC2_RAT	Q96W30 rattus norv
4	166.5	3.4	1429	1 L112_CAEEL	P14585 caenorhabdi
5	165	3.4	2531	1 NTC1_RAT	Q07008 rattus norv
6	163	3.3	2531	1 NTC1_MOUSE	Q01705 mus musculu
7	162.5	3.3	2437	1 NTC1_BRARE	P46530 brachydantio
8	154.5	3.1	2321	1 NTC4_HUMAN	Q9um47 homo sapien
9	152	3.1	1664	1 NTC4_MOUSE	P31695 mus musculu
10	151.5	3.1	2003	1 NTC4_HUMAN	Q99466 homo sapien
11	149.5	3.0	2224	1 NTC4_XENLA	P21783 xenopus lae
12	148	3.0	2319	1 NTC3_RAT	Q9r172 rattus norv
13	147	3.0	2318	1 NTC3_MOUSE	Q61982 mus musculu
14	145.5	3.0	2556	1 NTC1_HUMAN	P46531 homo sapien
15	145	3.0	626	1 GIDA_BUCBP	P59485 buchiera ap
16	137.5	2.8	1328	1 YMOO_YEAST	Q04670 saccharomyc
17	136.5	2.8	1701	1 MSP1_YEAST	P06869 plasmodium
18	135.5	2.8	1701	1 MSP1_PLAFW	P18819 plasmodium
19	134.5	2.7	2434	1 YCP1_OENHO	Q9mth5 oenothera h
20	134	2.7	1295	1 GLP1_CAEEL	P13508 caenorhabdi
21	131	2.7	2095	1 RPI_MOUSE	P56716 mus musculu
22	129.5	2.6	1630	1 MSP1_PLAFW	P04932 plasmodium
23	129.5	2.6	1639	1 MSP1_PLAFW	P04933 plasmodium
24	129.5	2.6	2867	1 RBP2_PLAVB	Q00799 plasmodium
25	129	2.6	2703	1 NTC2_DROME	P07207 drosophila
26	128.5	2.6	756	1 Y328_WYCGE	Q49419 mycoplasma
27	128	2.6	631	1 G12A_HUCAG	O51879 buchiera ap
28	127.5	2.6	780	1 MUS2_BORBU	O51125 buchiera bu
29	127	2.6	1459	1 GEAT_YEAST	P39993 saccharomyc
30	126.5	2.6	1002	1 CLNN_HUMAN	Q96192 homo sapien
31	126.5	2.6	1328	1 YMT5_YEAST	Q04214 saccharomyc
32	125.5	2.6	2738	1 PGCV_RAT	Q9eab4 rattus norv
33	124	2.5	3135	1 S230_PLAFO	Q06872 plasmodium

ALIGNMENTS

RESULT 1	NTC2_MOUSE	STANDARD;	PRT; 2470 AA.
ID	NTC2_MOUSE	O35516; Q06008; Q60941;	
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch B)		
GN	NOTCH2		
OS	Mus musculus (Mouse)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6; TISSUE=Thymus;		
RA	Hamada Y., Higuchi M., Tsujimoto Y.;		
RT	"Complete amino acid sequence and multicistronic transcripts encoded by a single copy of mouse Notch2 gene."		
RL	Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE OF 316-1518 FROM N.A.		
RC	STRAIN=C57BL/6 X CBA; TISSUE=Embryo;		
RX	MEDLINE=93178563; PubMed=8440332;		
RA	Lardelli M., Lendahl U.;		
RT	"Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety of tissues."		
RL	Exp. Cell Res. 204:364-372 (1993).		
RN	[3]		
RP	SEQUENCE OF 1765-2153 FROM N.A.		
RX	MEDLINE=97075110; PubMed=8917536;		
RA	Malner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D., Martin D.I.;		
RT	"Inhibition of granulocytic differentiation by mNotch1."		
RL	Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019 (1996).		
RN	[4]		
RP	FUNCTION		
RX	MEDLINE=99396706; PubMed=10393120;		
RA	Hamada Y., Kadohara Y., Okabe M., Ikawa M., Coleman J.R., Tsujimoto Y.;		
RT	"Mutation in ankyrin repeats of the mouse Notch2 gene induces early embryonic lethality."		
RL	Development 126:3415-3424 (1999).		
RN	[5]		
RP	DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.		
RX	MEDLINE=95333893; PubMed=7605614;		
RA	Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;		
RT	"Differential expression of Notch1 and Notch2 in developing and adult mouse brain."		
RL	Brain Res. Mol. Brain Res. 29:263-272 (1995).		
RN	[6]		
RP	POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.		
RX	MEDLINE=21523956; PubMed=11518718;		
RA	Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;		
RT	"Murine notch homologs (N1-4) undergo presenilin-dependent		

34	124	2.5	3259	1	G031_HUMAN	Q14789 homo sapien
35	123	2.5	770	1	DBP4_YEAST	P20448 saccharomyc
36	123	2.5	5430	1	MACP_HUMAN	Q9ubn3 homo sapien
37	122.5	2.5	1111	1	EXSC_BUCBP	Q98ab4 buchiera ap
38	122.5	2.5	1465	1	SPB2_YEAST	P23201 saccharomyc
39	122.5	2.5	1541	1	YH46_YEAST	P38873 saccharomyc
40	122.5	2.5	2763	1	TEBG_VZVD	P09278 varicella-z
41	122	2.5	1312	1	RA50_YEAST	P12753 saccharomyc
42	122	2.5	5938	1	MAC4_HUMAN	Q96pk2 homo sapien
43	121.5	2.5	1755	1	YJ29_YEAST	P47100 saccharomyc
44	121.5	2.5	2025	1	TTG3_HUMAN	P53804 homo sapien
45	121	2.5	1052	1	CLNN_MOUSE	Q8c5w0 mus musculu

[illegible]

```

Query Match      3.4%; Score 168.5; DB 1; Length 2470;
Best Local Similarity 27.4%; Pred. No. 0.042;
Matches 46; Conservative 12; Mismatches 49; Indels 61; Gaps 6;

OY 422 SHSKGQKYLTPVPCNCAGCGSWIKDGYCDKACNMGACPDWGDGSGNGSGGRYIAG 481
DB 1407 SHCELYTAPTSTPTTCOSQYCADARBGICDEACNSHACQWDGDC----- 1453
OY 482 GGTGGIGVGHWPQFGGINSVS-----YCNQGC-----AN 511
DB 1454 -----SLTMEDEPW-----ANCTSLRCWEYINNQCDEQCNTABCLPDEFCQNSKTKYD 1504
OY 512 SWLADKF-----CDQACNVLSCGFADGCGDGFHELYK-----VILLP 550
DB 1505 KICADHFQDNHCDQGNSEBEGCGWBDGLDQASDPENLAEGLIIVLLP 1552

RESULT 2
ID NT02 HUMAN STANDARD; PRT; 2471 AA.
AC 004721; Q99734; Q9H240;
DT 28-FEB-2003 (Rel. 41; Created)
DT 28-FEB-2003 (Rel. 41; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Neogenic locus notch homolog protein 2 precursor (Notch 2) (hN2).
GN NOTCH2.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RT Blaumweller C.M., Mann R.S.;
RL "Complete human notch 2 (hN2) cDNA sequence.";
RN [2]
RP Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RT Correa R.G., Camargo A.A., Moreira E.S., Simpson A.J.G.;
RL "Human Notch2, a novel member of cell-fate determining NOTCH
family.";
RN [4]
RP Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 967-1229 FROM N.A.
RC TISSUE=T-cell;
RT Lemasson I., Devaux C., Mesnard J.M.;
RL "Partial sequence of EGF-like repeat domain of human Notch2 mRNA.";
RN [6]
RP Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 1810-2447 FROM N.A.
RC TISSUE=Brain;
RT MEDLINE=93365135; PubMed=1303260;
RN [8]
RP MEDLINE=93365135; PubMed=1303260;
RN [9]
RP Stefani S., Blaumweller C.W., Redhead N.J., Hill R.E.,
RN [10]
RP Artavanis-Tsakonas S.;
RN [11]
RP "Human homologs of a Drosophila enhancer of split gene product define
a novel family of nuclear proteins.";
RN [12]
RP Nat. Genet. 2:119-127 (1992).
RN [13]
RP POST-TRANSLATIONAL PROCESSING.
RN [14]
RP MEDLINE=97386453; PubMed=9244302;
RN [15]
RP Blaumweller C.W., Qi H., Zagouras P., Artavanis-Tsakonas S.;
RN [16]
RP "Intracellular cleavage of Notch leads to a heterodimeric receptor on
the plasma membrane.";
RN [17]
RP Cell 90:281-291 (1997).
RN [18]
RP IDENTIFICATION OF LIGANDS.
RN [19]
RP MEDLINE=99180765; PubMed=10079256;
RN [20]
RP Gray G.B., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RN [21]
RP Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RN [22]
RP "Human ligands of the Notch receptor.";
RN [23]
RP Am. J. Pathol. 154:785-794 (1999).

```

```

CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity).
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- TISSUE SPECIFICITY: Expressed in the brain, heart, kidney, lung,
CC skeletal muscle and liver.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -1- PTM: Phosphorylated (By similarity).
CC -1- SIMILARITY: Contains 35 EGF-like domains.
CC -1- SIMILARITY: Contains 2 lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF308601; AAA36377.2; -.
DR EMBL; AF315356; AAG37073.1; -.
DR EMBL; U77493; AAB19224.1; -.
DR HSSP; P00740; 1EDM.
DR Genew; HGNC:7882; NOTCH2.
DR MIM; 600275; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PR00023; ank; 6.
DR Pfam; PR00008; EGF_35.
DR Pfam; PR00066; notch; 2.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFBLMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00004; NL; 2.
DR PROSITE; PSS0297; ANK_REPEAT; 1.
DR PROSITE; PSS0088; ANK_REPEAT; 4.
DR PROSITE; PSS0010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.
DR PROSITE; PS01187; EGF_3; 35.
DR PROSITE; PS01187; EGF_CA; 22.
DR PROSITE; Transcription regulation; Activator; Differentiation;
DR Receptor; Transcription regulation; Activator; Differentiation;
DR Developmental protein; Repeat; ANK repeat; EGF-like domain;

```

KW	Transmembrane, Glycoprotein, Signal, Phosphorylation.
FT	SIGNAL 1 25
FT	CHAIN 26 2471
FT	CHAIN 1666 2471
FT	CHAIN 1697 2471
FT	DOMAIN 26 1677
FT	TRANSMEM 1678 1698
FT	DOMAIN 1699 2471
FT	DOMAIN 26 63
FT	DOMAIN 64 102
FT	DOMAIN 105 143
FT	DOMAIN 144 180
FT	DOMAIN 182 219
FT	DOMAIN 221 258
FT	DOMAIN 260 296
FT	DOMAIN 298 336
FT	DOMAIN 338 374
FT	DOMAIN 375 413
FT	DOMAIN 415 454
FT	DOMAIN 456 492
FT	DOMAIN 494 530
FT	DOMAIN 532 568
FT	DOMAIN 570 605
FT	DOMAIN 607 643
FT	DOMAIN 645 680
FT	DOMAIN 682 718
FT	DOMAIN 720 755
FT	DOMAIN 757 793
FT	DOMAIN 831 867
FT	DOMAIN 873 909
FT	DOMAIN 911 947
FT	DOMAIN 949 985
FT	DOMAIN 1023 1059
FT	DOMAIN 1065 1101
FT	DOMAIN 1103 1147
FT	DOMAIN 1149 1185
FT	DOMAIN 1187 1223
FT	DOMAIN 1225 1262
FT	DOMAIN 1302 1334
FT	DOMAIN 1340 1374
FT	DOMAIN 1412 1446
FT	REPEAT 1420 1456
FT	REPEAT 1503 1535
FT	REPEAT 1827 1871
FT	REPEAT 1876 1905
FT	REPEAT 1909 1939
FT	REPEAT 1943 1972
FT	REPEAT 1976 2005
FT	REPEAT 2009 2038
FT	DOMAIN 1645 1648
FT	DOMAIN 1994 1997
FT	DOMAIN 2429 2429
FT	DISULFID 28 41
FT	DISULFID 35 51
FT	DISULFID 53 62
FT	DISULFID 68 79
FT	DISULFID 73 90
FT	DISULFID 92 101
FT	DISULFID 109 121
FT	DISULFID 115 131
FT	DISULFID 133 142
FT	DISULFID 148 159
FT	DISULFID 153 168
FT	DISULFID 170 179
FT	DISULFID 186 198
FT	DISULFID 192 207
FT	DISULFID 209 218
FT	DISULFID 225 236
FT	DISULFID 230 246
FT	DISULFID 230 246

FT	DISULFID 248 257	BY SIMILARITY.
Query Match 3.4%; Score 168.5; DB 1; Length 2471;		
Best Local Similarity 19.7%; Pred. No. 0.042;		
Matches 114; Conservative 66; Mismatches 185; Indels 21; Gaps 27;		
QY	434 PVNCAEGCGGSGWIKDYCDKACNNSACWDCGDCSGSGGSRYYIAGGGGTSGIGVGHF 493	
DB	1421 PPAATLSQYCADARARQGVDCACNSHACQMDGDC-----SLTWENPM 1463	
QY	494 -QGCGGINSVYCNOC-----ANS-----MLADKF-----CDQAN 524	
DB	1464 ANCSSPLPCMDYINNOCDELCTVBCLEPNECQGSCTCKYDKCADHFKONHCGN 1523	
QY	525 VLSGCPDAGDCQDHFHELYK-----VILLPNQTHIIPKGBCLPYFSPAEV----- 571	
DB	1524 SEECGMDGDCADQDQENLABEGLVIVLMF-----PEQLDQARSFLALGLTLHTN 1576	
QY	572 --AKRVEGASNPPIIRASIANKKKTHILMHSNATTHFNULTPONTNDEEFK-- 626	
DB	1577 LRIKRDQGBLWYFPYGEKSAAMKQ-----RMTRSLSGEGOEQEVA 1619	
QY	627 -MOITVEVDTREGPKLNTSAQGYENVLPITLPEAEILFEDIPEKRPFKKHVNS 685	
DB	1620 GSKVFELEIDNQ---CVQSDHCFKNTDAALHLS-----HALQG 1657	
QY	686 TRPAQEEVKIPLYNI---SLPRDAQSLNTLDLQ-----EHGDT 724	
DB	1658 T-----LSYPLVSVSESLTPEPTQL--LYLLAVAVIILFILLGVMAKRRKRGSLW 1710	
QY	725 L-KGYVLSKALRSPLANSQHAQ-----IKNOAITDENDSIVAQEKVKS 773	
DB	1711 LPEGFTLRDA-----SNKRREPVGQDVAGLKNSVQSEAN--LI----- 1750	
QY	774 ILPNSLVSERLQRLTFPAVSXKNGHDG---QNPDLLETARFVETHTKTIGN 829	
DB	1751 ---GTSSEHWVDEGPQPK-KVKAEDBALISEDDPIDR-----RPVQOHLBAA 1797	
QY	830 VIKKEKPSLIVLESQMTKEKKITGKEKENSMEENAHNGVTEVILGRKLQHTDYL 889	
DB	1798 DIRRPSLALTPPOA-----EQEVVDVAVVRGPDGCTYMLTA----- 1835	
QY	890 GFLPWEKKRYFODLDEESLKTOLA-YETDSKNVGRQK 928	
DB	1836 -----SLRGSSDLSDDEDADSSANITTDLVYQASLDQ 1870	
RESULT 3		
NTC2_RAT	STANDARD; PRT; 2471 AA.	
ID	NTC2_RAT	
AC	Q9QW30;	
DT	28-FEB-2003 (Rel. 41, Created)	
DT	28-FEB-2003 (Rel. 41, Last sequence update)	
DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	Neurogenic locus notch homolog protein 2 precursor (Notch 2).	
GN	NOTCH2.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RC	SEQUENCE FROM N.A.	
RP	TISSUE=Brain;	
RX	MEDLINE=93202015; PubMed=1295745;	
RA	Weinmaster G., Roberts V.J., Lemke G.;	
RT	"Notch2: a second mammalian Notch gene."	
RL	Development 116:931-941(1992).	
RN	[2]	
RP	TISSUE SPECIFICITY.	
RX	MEDLINE=21331789; PubMed=11438922;	
RA	Irvin D.K., Zucher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;	
RT	"Expression patterns of Notch1, Notch2, and Notch3 suggest multiple	
RT	functional roles for the Notch-Delta signaling system during brain	

RA Wilson J., Rainscough R., Anderson K., Baynes C., Berts M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Lareille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
 RA Sims W., Smaldon N., Smith A., Smith M., Sonnhammer E., Staeden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Spratt J.,
 RA Wooldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL elegans.";
 RL Nature 368:32-38(1994).
 CC -I- FUNCTION: Involved in several cell fates decisions that requires
 CC cell-cell interactions. It is possible that lin-12 encodes a
 CC membrane-bound receptor for a signal that enables expression of
 CC the ventral uterine precursor cell fate.
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -I- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.
 CC -I- SIMILARITY: Contains 13 EGF-like domains.
 CC -I- SIMILARITY: Contains 3 lin/Notch repeats.
 CC -I- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation on
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC	EMBL	M12069	AAA70191.1	;-
DR	EMBL	Z14092	CAA78474.1	;-
DR	PIR	S06434	S06434	
DR	HSSP	P00740	IEDM	
DR	MorpPed	R107.8	CE00274	
DR	InterPro	IPR002110	ANK	
DR	InterPro	IPR000152	Asx_hydroxy1_s	
DR	InterPro	IPR000742	EGF_2	
DR	InterPro	IPR001881	EGF_Ca	
DR	InterPro	IPR006209	EGF_like	
DR	InterPro	IPR020439	Laminin_EGF	
DR	InterPro	IPR008000	Notch_dom	
DR	Pfam	PF00023	ank	6
DR	Pfam	PF00008	EGF	13
DR	Pfam	PF00066	notch	3
DR	PRINTS	PRO0011	EGFLAMININ	
DR	PRINTS	PRO1452	NOTCH	
DR	SMART	SM00248	ANK	6
DR	SMART	SM00179	EGF_CA	2
DR	SMART	SM00004	NL	2
DR	PROSITE	PS50297	ANK_REPEAT	1
DR	PROSITE	PS50088	ANK_REPEAT	3
DR	PROSITE	PS00010	ASX_HYDROXYL	3
DR	PROSITE	PS00022	EGF_1	12
DR	PROSITE	PS01186	EGF_2	11
DR	PROSITE	PS50026	EGF_3	13
DR	PROSITE	PS01187	EGF_CA	2
KW	Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal.			
FT	SIGNAL	1	15	POTENTIAL.
FT	DOMAIN	16	1425	LIN-12 PROTEIN.
FT	DOMAIN	16	908	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	909	931	POTENTIAL.
FT	DOMAIN	932	1429	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	20	61	EGF-LIKE 1.
FT	DOMAIN	114	150	EGF-LIKE 2.
FT	DOMAIN	152	190	EGF-LIKE 3.
FT	DOMAIN	201	246	EGF-LIKE 4.
FT	DOMAIN	250	285	EGF-LIKE 5.
FT	DOMAIN	287	323	EGF-LIKE 6.
FT	DOMAIN	323	363	EGF-LIKE 7.

```

FT DOMAIN 365 402 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 404 441 EGF-LIKE 9.
FT DOMAIN 449 492 EGF-LIKE 10.
FT DOMAIN 503 541 EGF-LIKE 11.
FT DOMAIN 543 579 EGF-LIKE 12.
FT DOMAIN 582 619 EGF-LIKE 13.
FT REPEAT 635 669 LIN/NOTCH 1.
FT REPEAT 670 710 LIN/NOTCH 2.
FT REPEAT 711 750 LIN/NOTCH 3.
FT REPEAT 1093 1122 ANK 1.
FT REPEAT 1126 1158 ANK 2.
FT REPEAT 1162 1194 ANK 3.
FT REPEAT 1206 1236 ANK 4.
FT REPEAT 1240 1269 ANK 5.
FT DISULFID 24 35 BY SIMILARITY.
FT DISULFID 29 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 118 129 BY SIMILARITY.
FT DISULFID 123 138 BY SIMILARITY.
FT DISULFID 140 149 BY SIMILARITY.
FT DISULFID 156 169 BY SIMILARITY.
FT DISULFID 163 178 BY SIMILARITY.
FT DISULFID 180 189 BY SIMILARITY.
FT DISULFID 205 237 BY SIMILARITY.
FT DISULFID 221 234 BY SIMILARITY.
FT DISULFID 236 245 BY SIMILARITY.
FT DISULFID 254 264 BY SIMILARITY.
FT DISULFID 259 273 BY SIMILARITY.
FT DISULFID 275 284 BY SIMILARITY.
FT DISULFID 291 302 BY SIMILARITY.
FT DISULFID 296 311 BY SIMILARITY.
FT DISULFID 313 322 BY SIMILARITY.
FT DISULFID 327 339 BY SIMILARITY.
FT DISULFID 334 351 BY SIMILARITY.
FT DISULFID 353 362 BY SIMILARITY.
FT DISULFID 369 381 BY SIMILARITY.
FT DISULFID 375 390 BY SIMILARITY.
FT DISULFID 392 401 BY SIMILARITY.
FT DISULFID 408 419 BY SIMILARITY.
FT DISULFID 413 429 BY SIMILARITY.
FT DISULFID 431 440 BY SIMILARITY.
FT DISULFID 507 518 BY SIMILARITY.
FT DISULFID 512 529 BY SIMILARITY.
FT DISULFID 531 540 BY SIMILARITY.
FT DISULFID 547 558 BY SIMILARITY.
FT DISULFID 567 578 BY SIMILARITY.
FT DISULFID 586 597 BY SIMILARITY.
FT DISULFID 591 607 BY SIMILARITY.
FT DISULFID 609 618 BY SIMILARITY.
FT CARBOHYD 41 41 N-LINKED (GLCNA6. . .) (POTENTIAL).
FT CARBOHYD 155 165 N-LINKED (GLCNA6. . .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNA6. . .) (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNA6. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNA6. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNA6. . .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNA6. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNA6. . .) (POTENTIAL).
FT CARBOHYD 900 900 N-LINKED (GLCNA6. . .) (POTENTIAL).
SQ SEQUENCE 1429 AA; 157115 MW; 2558DDV7A62C035DB CRC64;

Query Match 3.4%; Score 166.5; DB 1; Length 1429;
Best Local Similarity 20.4%; Pred. No. 0.024;
Matches 113; Conservative 43; Mismatches 172; Indels 225; Gaps 22;

```

```

QY 529 GPDAGCGQDPHFEHYKVLIPNOTHYIIPKGBCLPYFSPFAEYAKRGVEGASDNPILRH 568
DB 743 GFDGSDC-----DNET----- 753
QY 589 ASIANKKWIKTHLMHSGMATTI-HPNLTPQNTNDEFK-----NOTIVEVD 634
DB 754 -----NATITNIRITVQ-MDPKEFCVTGGQSLMEISSALRVYRIQ 794
QY 635 -TRBEPKLNSTAQKVENLVSPITL-----LPEALIFEDIPKPKPKFPHVNSTRRA 669
DB 795 RDEBGPV---FQWNGSEMDRVKNERQUTEQHVLISTISR-----KIKRSATNIGVVV 846
QY 690 QEEVKIPLVNIISLPDQALSTNTLDLEHGDITLKGVLN-----KSALLRSFL- 740
DB 847 YLEFGNCDPDKGLYVDAGSVDSISARLAKKIDSFGLPISALVAEPRKSNNNGFLS 906
QY 741 -----MNSCHAKIKNOAITTDET 758
DB 907 WNAILLIGAGCLIVWVVLMLGALPGNRTKRRRMINASVMPMEWNEKRNKNQSISSQ 966
QY 759 NDSIVAPQEKV-----HKSILPNSLGVSF-----RLORLTPPAVSXKVG 799
DB 967 HSLHLSYDGYIKQRNELOHYSLYENPQSYGNGNDPLGDFNHTNLIQITPEPESEPIKL 1026
QY 800 HDGQNPPLDLETTARFVE---THQKT---IGGNVTKEKPPSLIVPLESQMTREKKI 852
DB 1027 HTEAGSVAITEPIETRESVNIIDPRHNRVYLHWIASNSAEKSEDILV---HEAKECIA 1082
QY 853 TGKEXENSRMEEN 865
DB 1083 AGADVAMDCDEN 1095

RESULT 5
NTCL RAT STANDARD; PRT; 2531 AA.
AC Q07006;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Schwann cell;
RX MEDLINE=9211383; PubMed=1764995;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian
RT development 113:199-205 (1991).
RN [2]
RP REVISIONS TO 1652-1653.
RA Weinmaster G.;
RP Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RN FUNCTION.
RX MEDLINE=21094508; PubMed=1182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA Horio T.;
RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT neural progenitor cells to an astroglial fate.";
RN Neuron 29:45-55 (2001).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=93202015; PubMed=1295745;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "Notch2: a second mammalian Notch gene.";
RN Development 116:931-941 (1992).
RN [5]

```


RP TISSUE SPECIFICITY.
 RX MEDLINE:21331789; PubMed:11438922;
 RA Irvn D.K., Zürcher S.D., Nguyen T., Weimaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 functional roles for the Notch-DSL signaling system during brain
 development.";
 RT J. Comp. Neurol. 436:167-181 (2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus (By
 CC similarity).
 CC -1- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
 CC Expressed in postnatal central nervous system (CNS) germinal zones
 CC and, in early postnatal life, within numerous cells throughout the
 CC CNS. Found in both subventricular and ventricular germinal zones.
 CC -1- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
 CC days 12 and 14 and decrease rapidly to much lower levels in the
 CC adult.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 36 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X57405; CA40667.1; -
 DR HSSP: P00740; IEDM.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxy1_S.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR006209; EGF_II.
 DR InterPro: IPR002949; Laminin_EGF.
 DR InterPro: IPR008287; Notch.
 DR InterPro: IPR000800; Notch_dom.
 DR Pfam: PF00023; ank_6.
 DR Pfam: PF00066; notch_3.
 DR Pfam: PF00066; notch_3.
 DR PIRSF: PIRSF002279; Notch; 1.
 DR PRINTS: PR00010; EGFLOD.
 DR PRINTS: PR00011; EGFAMININ.
 DR PRINTS: PR01452; NOTCH.

DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NT; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS01186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS01187; EGF_Ca; 21.
 DR Receptor; Transcription regulator; Activator; Differentiation;
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT
 FT CHAIN 1744 2531
 FT
 FT DOMAIN 19 1723
 FT TRANSMEM 1724 1746
 FT DOMAIN 1747 2531
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 604
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 867
 FT DOMAIN 869 905
 FT DOMAIN 907 943
 FT DOMAIN 945 981
 FT DOMAIN 983 1019
 FT DOMAIN 1021 1057
 FT DOMAIN 1059 1095
 FT DOMAIN 1097 1145
 FT DOMAIN 1145 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1265
 FT DOMAIN 1267 1305
 FT DOMAIN 1307 1346
 FT DOMAIN 1348 1384
 FT DOMAIN 1387 1426
 FT DOMAIN 1445 1480
 FT REPEAT 1481 1522
 FT REPEAT 1523 1562
 FT REPEAT 1917 1946
 FT REPEAT 1950 1980
 FT REPEAT 1984 2013
 FT REPEAT 2017 2046
 FT REPEAT 2050 2079
 FT DOMAIN 1730 1733
 FT DOMAIN 1891 1894
 FT DOMAIN 2258 2261
 FT DOMAIN 2497 2500
 FT SITE 1654 1655
 FT DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NT; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS01186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS01187; EGF_Ca; 21.
 DR Receptor; Transcription regulator; Activator; Differentiation;
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT
 FT CHAIN 1744 2531
 FT
 FT DOMAIN 19 1723
 FT TRANSMEM 1724 1746
 FT DOMAIN 1747 2531
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 604
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 867
 FT DOMAIN 869 905
 FT DOMAIN 907 943
 FT DOMAIN 945 981
 FT DOMAIN 983 1019
 FT DOMAIN 1021 1057
 FT DOMAIN 1059 1095
 FT DOMAIN 1097 1145
 FT DOMAIN 1145 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1265
 FT DOMAIN 1267 1305
 FT DOMAIN 1307 1346
 FT DOMAIN 1348 1384
 FT DOMAIN 1387 1426
 FT DOMAIN 1445 1480
 FT REPEAT 1481 1522
 FT REPEAT 1523 1562
 FT REPEAT 1917 1946
 FT REPEAT 1950 1980
 FT REPEAT 1984 2013
 FT REPEAT 2017 2046
 FT REPEAT 2050 2079
 FT DOMAIN 1730 1733
 FT DOMAIN 1891 1894
 FT DOMAIN 2258 2261
 FT DOMAIN 2497 2500
 FT SITE 1654 1655
 FT DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NT; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS01186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS01187; EGF_Ca; 21.
 DR Receptor; Transcription regulator; Activator; Differentiation;
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT
 FT CHAIN 1744 2531
 FT
 FT DOMAIN 19 1723
 FT TRANSMEM 1724 1746
 FT DOMAIN 1747 2531
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 604
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 867
 FT DOMAIN 869 905
 FT DOMAIN 907 943
 FT DOMAIN 945 981
 FT DOMAIN 983 1019
 FT DOMAIN 1021 1057
 FT DOMAIN 1059 1095
 FT DOMAIN 1097 1145
 FT DOMAIN 1145 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1265
 FT DOMAIN 1267 1305
 FT DOMAIN 1307 1346
 FT DOMAIN 1348 1384
 FT DOMAIN 1387 1426
 FT DOMAIN 1445 1480
 FT REPEAT 1481 1522
 FT REPEAT 1523 1562
 FT REPEAT 1917 1946
 FT REPEAT 1950 1980
 FT REPEAT 1984 2013
 FT REPEAT 2017 2046
 FT REPEAT 2050 2079
 FT DOMAIN 1730 1733
 FT DOMAIN 1891 1894
 FT DOMAIN 2258 2261
 FT DOMAIN 2497 2500
 FT SITE 1654 1655
 FT DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NT; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS01186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS01187; EGF_Ca; 21.
 DR Receptor; Transcription regulator; Activator; Differentiation;
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT
 FT CHAIN 1744 2531
 FT
 FT DOMAIN 19 1723
 FT TRANSMEM 1724 1746
 FT DOMAIN 1747 2531
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 604
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 867
 FT DOMAIN 869 905
 FT DOMAIN 907 943
 FT DOMAIN 945 981
 FT DOMAIN 983 1019
 FT DOMAIN 1021 1057
 FT DOMAIN 1059 1095
 FT DOMAIN 1097 1145
 FT DOMAIN 1145 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1265
 FT DOMAIN 1267 1305
 FT DOMAIN 1307 1346
 FT DOMAIN 1348 1384
 FT DOMAIN 1387 1426
 FT DOMAIN 1445 1480
 FT REPEAT 1481 1522
 FT REPEAT 1523 1562
 FT REPEAT 1917 1946
 FT REPEAT 1950 1980
 FT REPEAT 1984 2013
 FT REPEAT 2017 2046
 FT REPEAT 2050 2079
 FT DOMAIN 1730 1733
 FT DOMAIN 1891 1894
 FT DOMAIN 2258 2261
 FT DOMAIN 2497 2500
 FT SITE 1654 1655
 FT DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NT; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS01186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS01187; EGF_Ca; 21.
 DR Receptor; Transcription regulator; Activator; Differentiation;
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT
 FT CHAIN 1744 2531
 FT
 FT DOMAIN 19 1723
 FT TRANSMEM 1724 1746
 FT DOMAIN 1747 2531
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 604
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 867
 FT DOMAIN 869 905
 FT DOMAIN 907 943
 FT DOMAIN 945 981
 FT DOMAIN 983 1019
 FT DOMAIN 1021 1057
 FT DOMAIN 1059 1095
 FT DOMAIN 1097 1145
 FT DOMAIN 1145 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1265
 FT DOMAIN 1267 1305
 FT DOMAIN 1307 1346
 FT DOMAIN 1348 1384
 FT DOMAIN 1387 1426
 FT DOMAIN 1445 1480
 FT REPEAT 1481 1522
 FT REPEAT 1523 1562
 FT REPEAT 1917 1946
 FT REPEAT 1950 1980
 FT REPEAT 1984 2013
 FT REPEAT 2017 2046
 FT REPEAT 2050 2079
 FT DOMAIN 1730 1733
 FT DOMAIN 1891 1894
 FT DOMAIN 2258 2261
 FT DOMAIN 2497 2500
 FT SITE 1654 1655
 FT DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NT; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS01186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS01187; EGF_Ca; 21.
 DR Receptor; Transcription regulator; Activator; Differentiation;
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT
 FT CHAIN 1744 2531
 FT
 FT DOMAIN 19 1723
 FT TRANSMEM 1724 1746
 FT DOMAIN 1747 2531
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 604
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 867
 FT DOMAIN 869 905
 FT DOMAIN 907 943
 FT DOMAIN 945 981
 FT DOMAIN 983 1019
 FT DOMAIN 1021 1057
 FT DOMAIN 1059 1095
 FT DOMAIN 1097 1145
 FT DOMAIN 1145 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1265
 FT DOMAIN 1267 1305
 FT DOMAIN 1307 1346
 FT DOMAIN 1348 1384
 FT DOMAIN 1387 1426
 FT DOMAIN 1445 1480
 FT REPEAT 1481 1522
 FT REPEAT 1523 1562
 FT REPEAT 1917 1946
 FT REPEAT 1950 1980
 FT REPEAT 1984 2013
 FT REPEAT 2017 2046
 FT REPEAT 2050 2079
 FT DOMAIN 1730 1733
 FT DOMAIN 1891 1894
 FT DOMAIN 2258 2261
 FT DOMAIN 2497 2500
 FT SITE 1654 1655
 FT DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NT; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS01186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS01187; EGF_Ca; 21.
 DR Receptor; Transcription regulator; Activator; Differentiation;
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT
 FT CHAIN 1744 2531
 FT
 FT DOMAIN 19 1723
 FT TRANSMEM 1724 1746
 FT DOMAIN 1747 2531
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 604
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 867
 FT DOMAIN 869 905
 FT DOMAIN 907 943
 FT DOMAIN 945 981
 FT DOMAIN 983 1019
 FT DOMAIN 1021 1057
 FT DOMAIN 1059 1095
 FT DOMAIN 1097 1145
 FT DOMAIN 1145 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1265
 FT DOMAIN 1267 1305
 FT DOMAIN 1307 1346
 FT DOMAIN 1348 1384
 FT DOMAIN 1387 1426
 FT DOMAIN 1445 1480
 FT REPEAT 1481 1522
 FT REPEAT 1523 1562
 FT REPEAT 1917 1946
 FT REPEAT 1950 1980
 FT REPEAT 1984 2013
 FT REPEAT 2017 2046
 FT REPEAT 2050 2079
 FT DOMAIN 1730 1733
 FT DOMAIN 1891 1894
 FT DOMAIN 2258 2261
 FT DOMAIN 2497 2500
 FT SITE 1654 1655
 FT DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NT; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS01186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS01187; EGF_Ca; 21.
 DR Receptor; Transcription regulator; Activator; Differentiation;
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT
 FT CHAIN 1744 2531
 FT
 FT DOMAIN 19 1723
 FT TRANSMEM 1724 1746
 FT DOMAIN 1747 2531
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 604
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 867
 FT DOMAIN 869 905
 FT DOMAIN 907 943
 FT DOMAIN 945 981
 FT DOMAIN 983 1019
 FT DOMAIN 1021 1057
 FT DOMAIN 1059 1095
 FT DOMAIN 1097 1145
 FT DOMAIN 1145 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1265
 FT DOMAIN 1267 1305
 FT DOMAIN 1307 1346
 FT DOMAIN 1348 1384
 FT DOMAIN 1387 1426
 FT DOMAIN 1445 1480
 FT REPEAT 1481 1522
 FT REPEAT 1523 1562
 FT REPEAT 1917 1946
 FT REPEAT 1950 1980
 FT REPEAT 1984 2013
 FT REPEAT 2017 2046
 FT REPEAT 2050 2079
 FT DOMAIN 1730 1733
 FT DOMAIN 1891 1894
 FT DOMAIN 2258 2261
 FT DOMAIN 2497 2500
 FT SITE 1654 1655
 FT DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NT; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS01186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS01187; EGF_Ca; 21.
 DR Receptor; Transcription regulator; Activator; Differentiation;
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT
 FT CHAIN 1744 2531
 FT
 FT DOMAIN 19 1723
 FT TRANSMEM 1724 1746
 FT DOMAIN 1747 2531
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 604
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 867
 FT DOMAIN 869 905
 FT DOMAIN 907 943
 FT DOMAIN 945 981
 FT DOMAIN 983 1019
 FT DOMAIN 1021 1057
 FT DOMAIN 1059 1095
 FT DOMAIN 1097 1145
 FT DOMAIN 1145 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1265
 FT DOMAIN 1267 1305
 FT DOMAIN 1307 1346
 FT DOMAIN 1348 1384
 FT DOMAIN 1387 1426
 FT DOMAIN 1445 1480
 FT REPEAT 1481 1522
 FT REPEAT 1523 1562
 FT REPEAT 1917 1946
 FT REPEAT 1950 1980
 FT REPEAT 1984 2013
 FT REPEAT 2017 2046
 FT REPEAT 2050 2079
 FT DOMAIN 1730 1733
 FT DOMAIN 1891 1894
 FT DOMAIN 2258 2261
 FT DOMAIN 2497 2500
 FT SITE 1654 1655
 FT DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NT; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS01186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS01187; EGF_Ca; 21.
 DR Receptor; Transcription regulator; Activator; Differentiation;
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT
 FT CHAIN 1744 2531
 FT
 FT DOMAIN

```

FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.

Query Match 3.4%; Score 165; DB 1; Length 2531;
Best Local Similarity 20.6%; Pred.No.0.072; Mismatches 179; Indels 186; Gaps 27;
Matches 111; Conservative 63;

QY 412 GQVVPDDFYSHSKQKVTLPVPCNCAEGCPGSMIKDGYCDKACNNSACDWDGDCSGN 471
DB 1437 GRDIP-----PQIEACELPECCEDAGNK-----YCNLCNNHACGMDGDCSLN 1462
QY 472 SGGS-----RYLAG-----GGTSGTGVGHFWQGGGINSVSYCN---QCCA 510
DB 1483 FNDPMKNCTQSLQCKYFSDGHGDCSCNSAGCLPFGDCQLTEG-----QCNPLYDYCK 1537
QY 511 NSWLADKFCDOACNVLSGFGDAGCGDHFHE-----LYKYLTP-----NOTHYIIP 558
DB 1538 DHF-SDGHCDQCNSECEGMDGDCA-EHVEPRLAGLTVLVLPDQLRNNSHFIRLD 1535
QY 559 KKECL-----PYSPFAVAVRGVGAVSNDPITIRASLANKKTHLI 601
DB 1596 VSHVHTNVVPRDAQGOQMIIPY-----GREELRKPIRSAY--GWTATSL 1644
QY 602 -----MHSGMATTIHNLTFQNTNDEFPKQITVENPRE-----GP 639
DB 1645 PGTNGGRQRELDPMDDHSGIIV-----LEIDNQCVQSSQCFQATDVAA 1691
QY 640 KUNSTAQKGYENLVSPITLLPEAEILFEDIP-----KE 672
DB 1692 FLGALASIGSLNIPKIEAV-KSEIYEPPLPSQLHMTVAAPVLPFGVGYLSKR 1750
QY 673 KR-----FPK-FKHADVSTRAGBEVKIP--VNISLP-KDQSLNLTLDQLERG- 721
DB 1751 RSOHQQLWPEEGFKVSEAKKRR-----PLGEDSVGLKPLKNSDGLMDNDQNEWD 1805
QY 722 -DITLKYVLSKSLASFLMNSOHAQIKNOAIITDENDSLVAPOEQVHKSILPNSLG 780
DB 1806 EDELETKKFRFEEPVLPDLDDPTDRQWQHLDADLRVSMATPPQ----- 1854
QY 781 VSERLQRLTPAVSVKVGNDQGNPDLLETTARFVETHTQKTIGANVTEKEPPLI 839
DB 1855 -----GEVDADCDVAVRGPD--GFTPLMIASCSGGLT-----GNSSEEDAPAVI 1900

RESULT 6
NTC1_MOUSE STANDARD; PRT; 2531 AA.
AC Q01705; Q06007; Q61905; Q990C2; Q9QW58; Q9ROX7;
AT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Notch A)
DE (MIM14) (p3900).
GN NOTCH1 OR NOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Embryo;
RX MEDLINE=93194170; PubMed=8449489;
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RA Copeland N.G., Gridley T.;
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT homolog of Drosophila Notch."
RL Genomics 15:259-264 (1993).
RN [2]
RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
RC STRAIN=CD-1; TISSUE=Embryo;
RX MEDLINE=93050801; PubMed=1426644;
RA Reaume A.G., Conlon R.A., Ziringibl R., Yamaguchi T.P., Rossant J.;
RT "Expression analysis of a Notch homologue in the mouse embryo."
RL Dev. Biol. 154:377-387 (1992).
RN [3]
RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
RC TISSUE=Embryo;
RX MEDLINE=93048835; PubMed=1425352;
RA Franco del Amo F., Smith D.B., Swiatek P.J., Gendron-Maguire M.,
RA Greenpan R.J., McMahon A.P., Gridley T.;
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
RT suggests an important role in early postimplantation mouse
RT development."
RL Development 115:737-744 (1992).
RN [4]
RP SEQUENCE OF 1161-1547 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
RT wide variety of tissues."
RL Exp. Cell Res. 204:364-372 (1993).
RN [5]
RP SEQUENCE OF 1659-1673 FROM N.A.
RX MEDLINE=99364499; PubMed=10437788;
RA Lee J.S., Ishimoto A., Yanagawa S.I.;
RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3."
RL FEBS Lett. 455:276-280 (1999).
RN [6]
RP SEQUENCE OF 1950-2201 FROM N.A.
RX MEDLINE=98029496; PubMed=9384671;
RA Messerle M., Folio M., Nehls M., Eggert H., Boehm T.;
RT "Dynamic changes in gene expression during in vitro differentiation of
RT mouse embryonic stem cells."
RL Cytokines Cell. Mol. Ther. 1:139-143 (1995).
RN [7]
RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
RX MUTAGENESIS OF 1651-ARG-ARG-1654.
RX MEDLINE=98318619; PubMed=9651148;
RA Logeat F., Bessia C., Brou C., Debail O., Jarrault S., Seidah N.G.,
RA Israel A.;
RT "The Notch1 receptor is cleaved constitutively by a furin-like
RT convertase."
RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112 (1998).
RN [8]
RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis."
RL J. Biol. Chem. 276:40268-40273 (2001).
RN [9]
RP POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).
RN [10]
RP INTERACTION WITH DTX1 AND DTX2.
RX MEDLINE=21123790; PubMed=11226752;

```

RA Kishi N., Tang Z., Maeda Y., Hirai A., No R., Ito M., Suzuki S.,
RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Tsakonas S.,
RA Okano H., Matsuno K.,
RT "Murine homologs of *deltex* define a novel gene family involved in
RT vertebrate Notch signaling and neurogenesis." ;
RT Int. J. Dev. Neurosci. 19:21-35 (2001)
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC tagged, jagged and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation. May be involved in mesoderm
CC development, somitogenesis and neurogenesis. Involved in the
CC maturation of both CD4+ and CD8+ cells in the thymus.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
CC terminal fragment N(IEC) which are probably linked by disulfide
CC bonds. Interacts with DTX1 and DTX2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q01705-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
CC thymus. Expressed at lower levels in the spleen, bone-marrow,
CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
CC muscle, kidney and heart.
CC -1- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
CC 8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
CC endothelial cells, while much lower levels are seen in the
CC neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
CC the neuroepithelium. At 13.5 dpc expressed in the surface
CC ectoderm, eye and developing whisker follicles.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TW) and a N-terminal fragment N(IEC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 36 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 5 ANK repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; Z11886; CAA7941.1; -
CC EMBL; L02613; AAK14898.1; -
CC EMBL; X68278; CAA48339.1; -
CC EMBL; AJ238029; CAB40733.1; -
CC EMBL; X82562; CAA57909.1; -
CC PIR; A46019; A46019.
CC PIR; B49175; B49175.
CC HSSP; P00740; 1EDM.
CC MGI; MGI:97363; Notch1.

DR GO; GO:0005887; C: integral to plasma membrane; IC.
DR GO; GO:0005515; F: protein binding; IP.
DR GO; GO:0030154; P: cell differentiation; IMP.
DR GO; GO:0007386; P: compartment specification; IMP.
DR GO; GO:0007219; P: P signal pathway; IC.
DR GO; GO:0045944; P: positive regulation of transcription from P...; IDA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR008800; Notch_dom.
DR Pfam; PF00023; ank; 7.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF02279; Notch; 1.
DR PRINTS; PRO0010; EGFBLDOD.
DR PRINTS; PRO1452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_Ca; 24.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS50026; EGF_3; 36.
DR PROSITE; PS01187; EGF_Ca; 21.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT CHAIN 1711 2531 NOTCH EXTRACELLULAR TRUNCATION.
FT CHAIN 1744 2531 NOTCH INTRACELLULAR DOMAIN.
FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).
Query Match 3.3%; Score 163; DB 1; Length 2531;
Best Local Similarity 20.4%; Pred. No. 0.096;
Matches 119; Conservative 68; Mismatches 204; Indels 192; Gaps 29;
QY 434 PVNCAEGC-PGSMIKDG--YCDKACNNACDMDGDCSGSGS-----RYI 478
DB 1441 PPQIEBACELPEGCYDADENKVCNLCNNHACGMGDCGSLNFPNPKXCTGSLCKRYF 1500
QY 479 AGG-----GGTSGIVGHPWQFGGINSVSYCN-----QGANSLADKFCDOACNVLSG 528
DB 1501 SDHCSDQCNAGCLFDGFCDLTEG-----OCNPLYDQYCKDHF-SDGHCDCGNSABC 1554
QY 529 GFDAGDCGDHFE-----LYKVILP-----NOTHYIIRKGLCPYFSPAAYAKGV 576
DB 1555 EWDGLDCA-BHPERLAAGTLVAVLPDQLRNNSFHL---RELSHVLTNNVFKKDA 1610
QY 577 EGAISNPPIIRHNSAMK-----WKTILI-----MHGNNATTHFNLTFOPT 620
DB 1611 OGQGMIFPYGHEBELRKPIKRTVGWATSSLLPQSGQRRLDPMIDIGSIYV--- 1667
QY 621 NDEBFMQTVETDRE-----GPKLNSPAQGYNNVSPILLBAETL 665
DB 1668 -----LETDNQCVQSSQCFQATDVAAFLGALASLSNIPYKIAV-KSEPV 1716
QY 666 FEDIP-----KEKX-----FPK-FKHVDNSTRRAOE 691
DB 1717 EPPLEPQLHMYVAAAFVLPFLFVGQVLLSKRRRQHQLFPEPGFVSEASKKRR- 1775
QY 692 EVKIPV--VNISLP-KDAQLSLNTDLENG--DITKGYNTSKSALRFLNNSQRA 746
DB 1776 ---PLGSDSVGLKPLKNSDGLMDNDNGEWGDEDLTKKRFEBEPVLPDLSQDTQHR 1831

```

QY 747 KIKNGAIIIDETDSTLVAPEQKQVHKSILNSLGVSERLQRLTPFAVSQVKNHGQGNP 806
DB 1832 QWTOQHLDADADJMSMAFTPPQ-----GEVDADQMDVNVAGPD--GFT 1873
QY 807 PDLDETARFRVETHQKTIQGNVTKERKPSLI-----VLEQMTKEKKIT----- 853
DB 1874 PLMTASCSGGGLT-----GNSEEDAPAVISDFITYGASLHQTRTGRTGTLALHAR 1927
QY 854 -GKEKENSRENAENHI-----GTEVLL 877
DB 1928 YRSRDRKRRLBASADANIQDNNGRTPLHAASADAGVQIILL 1970

RESULT 7
NTCL BRARE STANDARD; PRT; 2437 AA.
AC P46530;
DT 01-NOV-1995 (Rel. 32, Last sequence created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor.
GN NOTCH1 OR NOTCH.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxId=7953;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94128602; PubMed=8297791;
RA Bierkamp C., Campos-Ortega J.A.;
RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RL its pattern of transcription during early embryogenesis.";
Meth. Dev. 43:87-100(1993).
CC - FUNCTION: Implicated in cell fate specifications during
CC embryonic development. May be involved in the formation of the
CC neural plate, notochord and brain vesicles.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation
CC stages. During gastrulation is differentially expressed,
CC accumulating predominantly in the prechordal mesoderm and
CC notochord. At the end of gastrulation, expressed along the
CC anterior-posterior axis including the developing neural plate
CC and differentiating mesoderm. Also present in the developing
CC brain and head regions.
CC - SIMILARITY: Belongs to the NOTCH family.
CC - SIMILARITY: Contains 36 EGF-like domains.
CC - SIMILARITY: Contains 3 Lin/Notch repeats.
CC - SIMILARITY: Contains 6 ANK repeats.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X69088; CAA48631.1; -.
DR F1R; S42612; S42612.
DR HSSP; P00740; 1EDM.
DR ZFIN; ZDB-GENE-990415-173; notch1a.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001981; EGF_Ca.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR PIRSF; PF00066; notch; 3.
DR PIRSF; PIRSF002379; Notch; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 22.
DR SMART; SM00004; NU; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 28.
DR PROSITE; PS50026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 22.
DR Receptor; Transcription regulation; Activator; Differentiation;
KM Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 785
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT DOMAIN 1446 1486
FT DOMAIN 1487 1520
FT DOMAIN 1521 1561
FT DOMAIN 1567 1594
FT DOMAIN 1595 1624
FT DOMAIN 1625 1656
FT DOMAIN 1657 1697
FT SITE 1656 1657
FT Pfam; PF00023; ank; 6.
FT Pfam; PF00008; EGF; 36.
FT PIRSF; PF00066; notch; 3.
FT PIRSF; PIRSF002379; Notch; 1.
FT PRINTS; PR00010; EGFBL00D.
FT PRINTS; PR00011; EGF_LAMININ.
FT PRINTS; PR01452; NOTCH.
FT SMART; SM00248; ANK; 6.
FT SMART; SM00179; EGF_CA; 22.
FT SMART; SM00004; NU; 3.
FT PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT PROSITE; PS50088; ANK_REPEAT; 4.
FT PROSITE; PS00010; ASX_HYDROXYL; 23.
FT PROSITE; PS00022; EGF_1; 34.
FT PROSITE; PS01186; EGF_2; 28.
FT PROSITE; PS50026; EGF_3; 36.
FT PROSITE; PS01187; EGF_CA; 22.
FT Receptor; Transcription regulation; Activator; Differentiation;
FT Developmental protein; Neurogenesis; Repeat; ANK repeat;
FT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 71
```



```

RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stiiwagen S.,
RA Phan H., Velasco N., Garmes J., Dangnan L., Poundstone P.,
RA Christensen M., Georgescu A., Avila J., Liu S., Atix C., Andreise T.,
RA Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
RA Carraro A.V.,
RT "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
RT 19p13.1-";
RL Submitted (May-1998) to the EMBL/Genbank/DBSD databases.
RN
RP VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
RP ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SER-212;
RP GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; ARG-578; CYS-728;
RP CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
RP ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
RX MEDLINE=98049753; PubMed=9388399,
RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,
RA Vaysiere C., Craud C., Maciazek J., Weissbach J., Bousser M.-G.,
RA Bach J.-F., Tournier-Lasserre E.,
RT "Strong clustering and stereotyped nature of Notch3 mutations in
RT CADASIL patients.";
RL Lancet 350:1511-1515(1997).
RN
RP VARIANTS CADASIL 114-GLY--PRO-120 DEL.
RX MEDLINE=20264473; PubMed=10802807;
RA Joutel A., Chabriat H., Vahedi K., Domenga V., Vaysiere C.,
RA Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserre E.,
RT "Splice site mutation causing a seven amino acid Notch3 in-frame
RT deletion in CADASIL.";
RL Neurology 54:1874-1875(2000).
RN
RN [6]
RP IDENTIFICATION OF LIGANDS.
RX MEDLINE=99180765; PubMed=10079256;
RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangu M.-L.,
RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Aravanis-Tsakonas S.,
RA "Human ligands of the Notch receptor.";
RL Am. J. Pathol. 154:785-794(1999).
RN
RN -1- FUNCTION: Functions as a receptor for membrane-bound ligands
RN Jagged1, Jagged2 and Delta to regulate cell-fate determination.
RN Upon ligand activation through the released notch intracellular
RN domain (NICD) it forms a transcriptional activator complex with
RN RBP-J kappa and activates genes of the enhancer of split locus.
RN Affects the implementation of differentiation, proliferation and
RN apoptotic programs (By similarity).
RN -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
RN terminal fragment N(IEC) which are probably linked by disulfide
RN bonds (By similarity).
RN -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
RN proteolytical processing NICD is translocated to the nucleus.
RN -1- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
RN tissues.
RN -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
RN which is proteolytically cleaved by a furin-like convertase in the
RN trans-golgi network before it reaches the plasma membrane to yield
RN an active, ligand-accessible form. Cleavage results in a C-
RN terminal fragment N(TW) and a N-terminal fragment N(IEC). Following
RN ligand binding, it is cleaved by TNF-alpha converting enzyme
RN (TACE) to yield a membrane-associated intermediate fragment called
RN notch extracellular truncation (NEXT). This fragment is then
RN cleaved by presenilin dependent gamma-secretase to release a
RN notch-derived peptide containing the intracellular domain (NICD)
RN from the membrane (By similarity).
RN -1- PTM: Phosphorylated (By similarity).
RN -1- DISORD: Defects in NOTCH3 are associated with cerebral autosomal
RN dominant arteriopathy with subcortical infarcts and
RN leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type
RN of stroke and dementia of which key features include recurrent
RN subcortical ischemic events and vascular dementia.
RN -1- SIMILARITY: Belongs to the NOTCH family.
RN -1- SIMILARITY: Contains 34 EGF-like domains.
RN -1- SIMILARITY: Contains 3 lin/Notch repeats.
RN -1- SIMILARITY: Contains 5 ANK repeats.
RN
RN -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, U97669; AAB91371.1; -.
DR EMBL, AF058900; AAC14346.1; -.
DR EMBL, AF058881; AAC14346.1; JOINED.
DR EMBL, AF058882; AAC14346.1; JOINED.
DR EMBL, AF058883; AAC14346.1; JOINED.
DR EMBL, AF058884; AAC14346.1; JOINED.
DR EMBL, AF058885; AAC14346.1; JOINED.
DR EMBL, AF058886; AAC14346.1; JOINED.
DR EMBL, AF058887; AAC14346.1; JOINED.
DR EMBL, AF058888; AAC14346.1; JOINED.
DR EMBL, AF058889; AAC14346.1; JOINED.
DR EMBL, AF058890; AAC14346.1; JOINED.
DR EMBL, AF058891; AAC14346.1; JOINED.
DR EMBL, AF058892; AAC14346.1; JOINED.
DR EMBL, AF058893; AAC14346.1; JOINED.
DR EMBL, AF058894; AAC14346.1; JOINED.
DR EMBL, AF058895; AAC14346.1; JOINED.
DR EMBL, AF058896; AAC14346.1; JOINED.
DR EMBL, AF058897; AAC14346.1; JOINED.
DR EMBL, AF058898; AAC14346.1; JOINED.
DR EMBL, AF058899; AAC14346.1; JOINED.
DR EMBL, AC004257; AAC04897.1; -.
DR EMBL, AC004653; AAC15789.1; ALT_INIT.
DR PIR, S78549; S78549.
DR HSPD, P00740; 1EDM.
DR Gensu; HGNC:7883; NOTCH3.
DR MIM, 600276; -.
DR MIM, 125310; -.
DR InterPro, IPR002110; ANK.
DR InterPro, IPR000152; Asx_hydroxyl_S.
DR InterPro, IPR000742; EGF_2.
DR InterPro, IPR001881; EGF_Ca.
DR InterPro, IPR001438; EGF_II.
DR InterPro, IPR006209; EGF_like.
DR InterPro, IPR002049; Laminin_EGF.
DR InterPro, IPR008297; Notch.
DR Pfam, PFO00800; Notch_dom.
DR Pfam, PFO0023; ank_6.
DR Pfam, PFO0006; notch_3.
DR Pfam, PFO0066; notch_1.
DR PRINTS, PRS002279; Notch_1.
DR PRINTS, PR00011; EGF_LAMININ.
DR PRINTS, PR001452; NOTCH.
DR SMART, SM00248; ANK_6.
DR SMART, SM00179; EGF_CA_19.
DR SMART, SM00004; NL_3.
DR PROSITE, PS50297; ANK_REPEAT_1.
DR PROSITE, PS50088; ANK_REPEAT_4.
DR PROSITE, PS00010; ASX_HYDROXYL_18.
DR PROSITE, PS00022; EGF_1_33.
DR PROSITE, PS01186; EGF_2_25.
DR PROSITE, PS50026; EGF_3_34.
DR PROSITE, PS01187; EGF_CA_16.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
KW Disease mutation.
FT SIGNAL 1 39
FT CHAIN 40 2321
FT CHAIN 1629 2321
FT CHAIN 1662 2321
FT CHAIN

```


FT DOMAIN 40 1643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1644 1664 POTENTIAL.
FT DOMAIN 1665 2321 CTOPOLASTIC (POTENTIAL).
FT DOMAIN 40 77 EGF-LIKE 1.
FT DOMAIN 78 118 EGF-LIKE 2.
FT DOMAIN 119 156 EGF-LIKE 3.
FT DOMAIN 158 195 EGF-LIKE 4.
FT DOMAIN 197 234 CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 236 272 EGF-LIKE 5.
FT DOMAIN 274 312 EGF-LIKE 6.
FT DOMAIN 314 350 EGF-LIKE 7.
FT DOMAIN 351 389 EGF-LIKE 8.
FT DOMAIN 391 429 EGF-LIKE 9.
FT DOMAIN 431 467 EGF-LIKE 10.
FT DOMAIN 469 505 EGF-LIKE 11.
FT DOMAIN 507 543 EGF-LIKE 12.
FT DOMAIN 545 580 EGF-LIKE 13.
FT DOMAIN 582 618 EGF-LIKE 14.
FT DOMAIN 620 655 EGF-LIKE 15.
FT DOMAIN 657 693 EGF-LIKE 16.
FT DOMAIN 695 730 EGF-LIKE 17.
FT DOMAIN 734 770 EGF-LIKE 18.
FT DOMAIN 774 808 EGF-LIKE 19.
FT DOMAIN 810 847 EGF-LIKE 20.
FT DOMAIN 849 885 EGF-LIKE 21.
FT DOMAIN 922 960 EGF-LIKE 22.
FT DOMAIN 962 1000 EGF-LIKE 23.
Query Match 3.1% Score 154.5; DB 1; Length 2321;
Best Local Similarity 24.7%; Pred No. 0.29; Indels 65; Gaps 6;
Matches 38; Conservative 14; Mismatches 37; Indels 65; Gaps 6;
QY 433 WPVNC-----AEGCPGSMK---DGYDKAKNSACDMGDCGSGSGSRY 477
DB 1367 WTGPRCEAPAAPEVSEBRCRACQARBPQRCDRECNBPGCGMDGDC----- 1417
QY 478 IAGGGGTGIGHGHPGFGGIN-----SVSYCNQGCAN----- 511
DB 1418 -----SLWGPDPGCEALQCKRLFNNSRCDPACSSPACLYNPDCHAGRERTNP 1469
QY 512 ---SWLADK-----CDQACNVLSGCFDAGDCGQD 538
DB 1470 VYKRYCADHPADGRCDQCGNTECGMDGDCASE 1503
RESULT 9
NTC4 MOUSE
ID NTC4 MOUSE STANDARD; PRT; 1964 AA.
AC P31655; O35442; O88314; O88316; Q62389; Q62390; Q9R1W5; Q9R1X0;
DT 01-JUL-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neuregulin locus notch homolog protein 4 precursor (Notch 4)
DB Contains: Transforming protein int-3.
GN NOTCH4 OR INT3 OR INT-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92194507; PubMed=1312643;
RA Robins J., Blondel B.J., Callahan D., Callahan R.;
RT "Mouse mammary tumor gene int-3: a member of the notch gene family
RT transforms mammary epithelial cells";
RL J. VIROL. 66:2594-2599 (1992).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RX MEDLINE=97294599; PubMed=9150355;
RA Callahan D., Callahan R.;
RT "The mouse mammary tumor associated gene INT3 is a unique member of
RT the NOTCH gene family (NOTCH4).";
RL Oncogene 14:1883-1890 (1997).
RN [3]

RP SEQUENCE FROM N.A.
RC TISSUE=Lung, and Testis;
RX MEDLINE=96281658; PubMed=8681805;
RA Uytendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT cell-specific mammalian Notch gene";
RL Development 122:2251-2259 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Rowen L., Nihalas G., Qin S., Ahearn M.E., Dankers C., Lasky S.;
RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class III
RT region";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1436-1600 FROM N.A.
RX MEDLINE=99252212; PubMed=10233992;
RA Lee J.-S., Harna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
RT "Intracutaneous type A particle-mediated activation of the Notch4/int3
RT gene in a mouse mammary tumor: generation of truncated Notch4/int3
RT mRNAs by retroviral splicing events";
RL J. VIROL. 73:5166-5171 (1999).
RN [6]
RP FUNCTION.
RX MEDLINE=21244657; PubMed=11344305;
RA Uytendaele H., Ho J., Rossant J., Kitajewski J.;
RT "Vascular patterning defects associated with expression of activated
RT Notch4 in embryonic endothelium";
RL Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648 (2001).
RN [7]
RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
OF VAL-1463.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis";
RL J. Biol. Chem. 276:40268-40273 (2001).
RN [8]
RP POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).
CC -I- FUNCTION: Functions as a receptor for membrane-bound ligands
Daggedl, Jagged2 and Delta to regulate cell-fate determination.
Upon ligand activation through the released notch intracellular
domain (NICD) it forms a transcriptional activator complex with
RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
apoptotic programs (by similarity). May regulate branching
morphogenesis in the developing vascular system.
CC -I- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
terminal fragment N(EC) which are probably linked by disulfide
bonds.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein. Following
proteolytic processing NICD is translocated to the nucleus.
CC -I- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
kidney, and at lower levels in the ovary and skeletal muscle. A
very low expression is seen in the brain, intestine, liver and
testis.
CC -I- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
embryonic development from 9.0 dpc.
CC -I- PTM: Synthesized in the endoplasmic reticulum as an inactive form
which is proteolytically cleaved by a furin-like convertase to yield
an active, ligand-accessible form. Cleavage results in a C-
terminal fragment N(TW) and a N-terminal fragment N(EC). Following
ligand binding, it is cleaved by TNF-alpha converting enzyme
(TACE) to yield a membrane-associated intermediate fragment called
notch extracellular truncation (NEXT). This fragment is then
cleaved by presenilin dependent gamma-secretase to release a
notch-derived peptide containing the intracellular domain (NICD)

CC	EMBL	D63395	BAA09708.1	ALT_FRAME.
CC	EMBL	D63395	BAA09708.1	ALT_FRAME.
DR	EMBL	D65565	BAA13116.1	-
DR	EMBL	D95299	AAC32288.1	-
DR	EMBL	U95353	AAC63097.1	-
DR	EMBL	AB023561	BAA20317.1	-
DR	EMBL	AB024520	BAA88951.1	-
DR	EMBL	AB024578	BAA88952.1	-
DR	HSSP	P08709	1BF9	-
DR	Genew	HGNC:7884	NOTCH4	-
DR	MTM	164851	-	-
DR	InterPro	IPR002110	ANK.	-
DR	InterPro	IPR000152	Asx_hydroxy1_S.	-
DR	InterPro	IPR000742	EGF_2.	-
DR	InterPro	IPR001881	EGF_Ca.	-
DR	InterPro	IPR001438	EGF_II.	-
DR	InterPro	IPR006209	EGF_like.	-
DR	InterPro	IPR002049	Laminin_EGF.	-
DR	InterPro	IPR008297	Notch.	-
DR	InterPro	IPR008000	Notch_dom.	-
DR	Pfam	PF00023	ank; 6.	-
DR	Pfam	PF00008	EGF; 26.	-
DR	Pfam	PF00066	notch; 2.	-
DR	PIRSP	PIRSP002379	Notch; 1.	-
DR	PRINTS	PR00010	EGFBLLOOD.	-
DR	PRINTS	PR00011	EGFLAMININ.	-
DR	SMART	SM00248	ANK; 5.	-
DR	SMART	SM00179	EGF_Ca; 11.	-
DR	SMART	SM00004	NL; 3.	-
DR	PROSITE	PS50297	ANK_REPEAT_1.	-
DR	PROSITE	PS50088	ANK_REPEAT_5.	-
DR	PROSITE	PS00010	ASX_HYDROXYL; 11.	-
DR	PROSITE	PS00022	EGF_1; 28.	-
DR	PROSITE	PS00186	EGF_2; 21.	-
DR	PROSITE	PS50026	EGF_3; 28.	-
DR	PROSITE	PS0187	EGF_CA; 9.	-
KV	Receptor	Transcription regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism; Triplet repeat expansion; Alternative splicing.	-	-
KW	Signal	1	23	POTENTIAL.
FT	CHAIN	24	2003	NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 4.
FT	CHAIN	1432	2003	NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY)
FT	CHAIN	1467	2003	NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY)
FT	DOMAIN	24	1447	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1448	1468	POTENTIAL.
FT	DOMAIN	1469	2003	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	24	63	EGF-LIKE 1.
FT	DOMAIN	64	115	EGF-LIKE 2.
FT	DOMAIN	118	155	EGF-LIKE 3.
FT	DOMAIN	156	192	EGF-LIKE 4.
FT	DOMAIN	194	232	EGF-LIKE 5.
FT	DOMAIN	234	274	EGF-LIKE 6.
FT	DOMAIN	276	312	EGF-LIKE 7.
FT	DOMAIN	314	353	EGF-LIKE 8.
FT	DOMAIN	355	391	EGF-LIKE 9.
FT	DOMAIN	392	430	EGF-LIKE 10.
FT	DOMAIN	432	473	EGF-LIKE 11.
FT	DOMAIN	475	511	EGF-LIKE 12.
FT	DOMAIN	513	549	EGF-LIKE 13.
FT	DOMAIN	551	587	EGF-LIKE 14.
FT	DOMAIN	589	625	EGF-LIKE 15.
FT	DOMAIN	626	659	EGF-LIKE 16.
FT	DOMAIN	661	689	EGF-LIKE 17.
FT	DOMAIN	691	727	EGF-LIKE 18.
FT	DOMAIN	729	765	EGF-LIKE 19.
FT	DOMAIN	767	803	EGF-LIKE 20.
FT	DOMAIN	806	842	EGF-LIKE 21.
FT	DOMAIN	844	880	EGF-LIKE 22.
FT	DOMAIN	882	928	EGF-LIKE 23.

```
FT DOMAIN 930 966 EGF-LIKE 24.
FT DOMAIN 968 1004 EGF-LIKE 25.
FT DOMAIN 1006 1044 EGF-LIKE 26.
FT DOMAIN 1046 1085 EGF-LIKE 27.
FT DOMAIN 1087 1126 EGF-LIKE 28.
FT DOMAIN 1130 1171 EGF-LIKE 29.
FT DOMAIN 1172 1212 POLY-ARG.
FT REPEAT 1213 1246 LTN/NOTCH 1.
FT REPEAT 1247 1286 LTN/NOTCH 2.
FT REPEAT 1287 1326 LTN/NOTCH 3.
FT REPEAT 1327 1366 ANK 1.
FT REPEAT 1367 1406 ANK 2.
FT REPEAT 1407 1446 ANK 3.
FT REPEAT 1447 1486 ANK 4.
FT REPEAT 1487 1526 ANK 5.
FT REPEAT 1527 1566 ANK 6.
FT REPEAT 1567 1606 ANK 7.
FT REPEAT 1607 1646 ANK 8.
FT REPEAT 1647 1686 ANK 9.
FT REPEAT 1687 1726 ANK 10.
FT REPEAT 1727 1766 ANK 11.
FT REPEAT 1767 1806 ANK 12.
FT REPEAT 1807 1846 ANK 13.
FT REPEAT 1847 1886 ANK 14.
FT REPEAT 1887 1926 ANK 15.
FT REPEAT 1927 1966 ANK 16.
FT REPEAT 1967 2006 ANK 17.
FT REPEAT 2007 2046 ANK 18.
FT REPEAT 2047 2086 ANK 19.
FT REPEAT 2087 2126 ANK 20.
FT REPEAT 2127 2166 ANK 21.
FT REPEAT 2167 2206 ANK 22.
FT REPEAT 2207 2246 ANK 23.
FT REPEAT 2247 2286 ANK 24.
FT REPEAT 2287 2326 ANK 25.
FT REPEAT 2327 2366 ANK 26.
FT REPEAT 2367 2406 ANK 27.
FT REPEAT 2407 2446 ANK 28.
FT REPEAT 2447 2486 ANK 29.
FT REPEAT 2487 2526 ANK 30.
FT REPEAT 2527 2566 ANK 31.
FT REPEAT 2567 2606 ANK 32.
FT REPEAT 2607 2646 ANK 33.
FT REPEAT 2647 2686 ANK 34.
FT REPEAT 2687 2726 ANK 35.
FT REPEAT 2727 2766 ANK 36.
FT REPEAT 2767 2806 ANK 37.
FT REPEAT 2807 2846 ANK 38.
FT REPEAT 2847 2886 ANK 39.
FT REPEAT 2887 2926 ANK 40.
FT REPEAT 2927 2966 ANK 41.
FT REPEAT 2967 3006 ANK 42.
FT REPEAT 3007 3046 ANK 43.
FT REPEAT 3047 3086 ANK 44.
FT REPEAT 3087 3126 ANK 45.
FT REPEAT 3127 3166 ANK 46.
FT REPEAT 3167 3206 ANK 47.
FT REPEAT 3207 3246 ANK 48.
FT REPEAT 3247 3286 ANK 49.
FT REPEAT 3287 3326 ANK 50.
FT REPEAT 3327 3366 ANK 51.
FT REPEAT 3367 3406 ANK 52.
FT REPEAT 3407 3446 ANK 53.
FT REPEAT 3447 3486 ANK 54.
FT REPEAT 3487 3526 ANK 55.
FT REPEAT 3527 3566 ANK 56.
FT REPEAT 3567 3606 ANK 57.
FT REPEAT 3607 3646 ANK 58.
FT REPEAT 3647 3686 ANK 59.
FT REPEAT 3687 3726 ANK 60.
FT REPEAT 3727 3766 ANK 61.
FT REPEAT 3767 3806 ANK 62.
FT REPEAT 3807 3846 ANK 63.
FT REPEAT 3847 3886 ANK 64.
FT REPEAT 3887 3926 ANK 65.
FT REPEAT 3927 3966 ANK 66.
FT REPEAT 3967 4006 ANK 67.
FT REPEAT 4007 4046 ANK 68.
FT REPEAT 4047 4086 ANK 69.
FT REPEAT 4087 4126 ANK 70.
FT REPEAT 4127 4166 ANK 71.
FT REPEAT 4167 4206 ANK 72.
FT REPEAT 4207 4246 ANK 73.
FT REPEAT 4247 4286 ANK 74.
FT REPEAT 4287 4326 ANK 75.
FT REPEAT 4327 4366 ANK 76.
FT REPEAT 4367 4406 ANK 77.
FT REPEAT 4407 4446 ANK 78.
FT REPEAT 4447 4486 ANK 79.
FT REPEAT 4487 4526 ANK 80.
FT REPEAT 4527 4566 ANK 81.
FT REPEAT 4567 4606 ANK 82.
FT REPEAT 4607 4646 ANK 83.
FT REPEAT 4647 4686 ANK 84.
FT REPEAT 4687 4726 ANK 85.
FT REPEAT 4727 4766 ANK 86.
FT REPEAT 4767 4806 ANK 87.
FT REPEAT 4807 4846 ANK 88.
FT REPEAT 4847 4886 ANK 89.
FT REPEAT 4887 4926 ANK 90.
FT REPEAT 4927 4966 ANK 91.
FT REPEAT 4967 5006 ANK 92.
FT REPEAT 5007 5046 ANK 93.
FT REPEAT 5047 5086 ANK 94.
FT REPEAT 5087 5126 ANK 95.
FT REPEAT 5127 5166 ANK 96.
FT REPEAT 5167 5206 ANK 97.
FT REPEAT 5207 5246 ANK 98.
FT REPEAT 5247 5286 ANK 99.
FT REPEAT 5287 5326 ANK 100.
FT REPEAT 5327 5366 ANK 101.
FT REPEAT 5367 5406 ANK 102.
FT REPEAT 5407 5446 ANK 103.
FT REPEAT 5447 5486 ANK 104.
FT REPEAT 5487 5526 ANK 105.
FT REPEAT 5527 5566 ANK 106.
FT REPEAT 5567 5606 ANK 107.
FT REPEAT 5607 5646 ANK 108.
FT REPEAT 5647 5686 ANK 109.
FT REPEAT 5687 5726 ANK 110.
FT REPEAT 5727 5766 ANK 111.
FT REPEAT 5767 5806 ANK 112.
FT REPEAT 5807 5846 ANK 113.
FT REPEAT 5847 5886 ANK 114.
FT REPEAT 5887 5926 ANK 115.
FT REPEAT 5927 5966 ANK 116.
FT REPEAT 5967 6006 ANK 117.
FT REPEAT 6007 6046 ANK 118.
FT REPEAT 6047 6086 ANK 119.
FT REPEAT 6087 6126 ANK 120.
FT REPEAT 6127 6166 ANK 121.
FT REPEAT 6167 6206 ANK 122.
FT REPEAT 6207 6246 ANK 123.
FT REPEAT 6247 6286 ANK 124.
FT REPEAT 6287 6326 ANK 125.
FT REPEAT 6327 6366 ANK 126.
FT REPEAT 6367 6406 ANK 127.
FT REPEAT 6407 6446 ANK 128.
FT REPEAT 6447 6486 ANK 129.
FT REPEAT 6487 6526 ANK 130.
FT REPEAT 6527 6566 ANK 131.
FT REPEAT 6567 6606 ANK 132.
FT REPEAT 6607 6646 ANK 133.
FT REPEAT 6647 6686 ANK 134.
FT REPEAT 6687 6726 ANK 135.
FT REPEAT 6727 6766 ANK 136.
FT REPEAT 6767 6806 ANK 137.
FT REPEAT 6807 6846 ANK 138.
FT REPEAT 6847 6886 ANK 139.
FT REPEAT 6887 6926 ANK 140.
FT REPEAT 6927 6966 ANK 141.
FT REPEAT 6967 7006 ANK 142.
FT REPEAT 7007 7046 ANK 143.
FT REPEAT 7047 7086 ANK 144.
FT REPEAT 7087 7126 ANK 145.
FT REPEAT 7127 7166 ANK 146.
FT REPEAT 7167 7206 ANK 147.
FT REPEAT 7207 7246 ANK 148.
FT REPEAT 7247 7286 ANK 149.
FT REPEAT 7287 7326 ANK 150.
FT REPEAT 7327 7366 ANK 151.
FT REPEAT 7367 7406 ANK 152.
FT REPEAT 7407 7446 ANK 153.
FT REPEAT 7447 7486 ANK 154.
FT REPEAT 7487 7526 ANK 155.
FT REPEAT 7527 7566 ANK 156.
FT REPEAT 7567 7606 ANK 157.
FT REPEAT 7607 7646 ANK 158.
FT REPEAT 7647 7686 ANK 159.
FT REPEAT 7687 7726 ANK 160.
FT REPEAT 7727 7766 ANK 161.
FT REPEAT 7767 7806 ANK 162.
FT REPEAT 7807 7846 ANK 163.
FT REPEAT 7847 7886 ANK 164.
FT REPEAT 7887 7926 ANK 165.
FT REPEAT 7927 7966 ANK 166.
FT REPEAT 7967 8006 ANK 167.
FT REPEAT 8007 8046 ANK 168.
FT REPEAT 8047 8086 ANK 169.
FT REPEAT 8087 8126 ANK 170.
FT REPEAT 8127 8166 ANK 171.
FT REPEAT 8167 8206 ANK 172.
FT REPEAT 8207 8246 ANK 173.
FT REPEAT 8247 8286 ANK 174.
FT REPEAT 8287 8326 ANK 175.
FT REPEAT 8327 8366 ANK 176.
FT REPEAT 8367 8406 ANK 177.
FT REPEAT 8407 8446 ANK 178.
FT REPEAT 8447 8486 ANK 179.
FT REPEAT 8487 8526 ANK 180.
FT REPEAT 8527 8566 ANK 181.
FT REPEAT 8567 8606 ANK 182.
FT REPEAT 8607 8646 ANK 183.
FT REPEAT 8647 8686 ANK 184.
FT REPEAT 8687 8726 ANK 185.
FT REPEAT 8727 8766 ANK 186.
FT REPEAT 8767 8806 ANK 187.
FT REPEAT 8807 8846 ANK 188.
FT REPEAT 8847 8886 ANK 189.
FT REPEAT 8887 8926 ANK 190.
FT REPEAT 8927 8966 ANK 191.
FT REPEAT 8967 9006 ANK 192.
FT REPEAT 9007 9046 ANK 193.
FT REPEAT 9047 9086 ANK 194.
FT REPEAT 9087 9126 ANK 195.
FT REPEAT 9127 9166 ANK 196.
FT REPEAT 9167 9206 ANK 197.
FT REPEAT 9207 9246 ANK 198.
FT REPEAT 9247 9286 ANK 199.
FT REPEAT 9287 9326 ANK 200.
FT REPEAT 9327 9366 ANK 201.
FT REPEAT 9367 9406 ANK 202.
FT REPEAT 9407 9446 ANK 203.
FT REPEAT 9447 9486 ANK 204.
FT REPEAT 9487 9526 ANK 205.
FT REPEAT 9527 9566 ANK 206.
FT REPEAT 9567 9606 ANK 207.
FT REPEAT 9607 9646 ANK 208.
FT REPEAT 9647 9686 ANK 209.
FT REPEAT 9687 9726 ANK 210.
FT REPEAT 9727 9766 ANK 211.
FT REPEAT 9767 9806 ANK 212.
FT REPEAT 9807 9846 ANK 213.
FT REPEAT 9847 9886 ANK 214.
FT REPEAT 9887 9926 ANK 215.
FT REPEAT 9927 9966 ANK 216.
FT REPEAT 9967 10006 ANK 217.
FT REPEAT 10007 10046 ANK 218.
FT REPEAT 10047 10086 ANK 219.
FT REPEAT 10087 10126 ANK 220.
FT REPEAT 10127 10166 ANK 221.
FT REPEAT 10167 10206 ANK 222.
FT REPEAT 10207 10246 ANK 223.
FT REPEAT 10247 10286 ANK 224.
FT REPEAT 10287 10326 ANK 225.
FT REPEAT 10327 10366 ANK 226.
FT REPEAT 10367 10406 ANK 227.
FT REPEAT 10407 10446 ANK 228.
FT REPEAT 10447 10486 ANK 229.
FT REPEAT 10487 10526 ANK 230.
FT REPEAT 10527 10566 ANK 231.
FT REPEAT 10567 10606 ANK 232.
FT REPEAT 10607 10646 ANK 233.
FT REPEAT 10647 10686 ANK 234.
FT REPEAT 10687 10726 ANK 235.
FT REPEAT 10727 10766 ANK 236.
FT REPEAT 10767 10806 ANK 237.
FT REPEAT 10807 10846 ANK 238.
FT REPEAT 10847 10886 ANK 239.
FT REPEAT 10887 10926 ANK 240.
FT REPEAT 10927 10966 ANK 241.
FT REPEAT 10967 11006 ANK 242.
FT REPEAT 11007 11046 ANK 243.
FT REPEAT 11047 11086 ANK 244.
FT REPEAT 11087 11126 ANK 245.
FT REPEAT 11127 11166 ANK 246.
FT REPEAT 11167 11206 ANK 247.
FT REPEAT 11207 11246 ANK 248.
FT REPEAT 11247 11286 ANK 249.
FT REPEAT 11287 11326 ANK 250.
FT REPEAT 11327 11366 ANK 251.
FT REPEAT 11367 11406 ANK 252.
FT REPEAT 11407 11446 ANK 253.
FT REPEAT 11447 11486 ANK 254.
FT REPEAT 11487 11526 ANK 255.
FT REPEAT 11527 11566 ANK 256.
FT REPEAT 11567 11606 ANK 257.
FT REPEAT 11607 11646 ANK 258.
FT REPEAT 11647 11686 ANK 259.
FT REPEAT 11687 11726 ANK 260.
FT REPEAT 11727 11766 ANK 261.
FT REPEAT 11767 11806 ANK 262.
FT REPEAT 11807 11846 ANK 263.
FT REPEAT 11847 11886 ANK 264.
FT REPEAT 11887 11926 ANK 265.
FT REPEAT 11927 11966 ANK 266.
FT REPEAT 11967 12006 ANK 267.
FT REPEAT 12007 12046 ANK 268.
FT REPEAT 12047 12086 ANK 269.
FT REPEAT 12087 12126 ANK 270.
FT REPEAT 12127 12166 ANK 271.
FT REPEAT 12167 12206 ANK 272.
FT REPEAT 12207 12246 ANK 273.
FT REPEAT 12247 12286 ANK 274.
FT REPEAT 12287 12326 ANK 275.
FT REPEAT 12327 12366 ANK 276.
FT REPEAT 12367 12406 ANK 277.
FT REPEAT 12407 12446 ANK 278.
FT REPEAT 12447 12486 ANK 279.
FT REPEAT 12487 12526 ANK 280.
FT REPEAT 12527 12566 ANK 281.
FT REPEAT 12567 12606 ANK 282.
FT REPEAT 12607 12646 ANK 283.
FT REPEAT 12647 12686 ANK 284.
FT REPEAT 12687 12726 ANK 285.
FT REPEAT 12727 12766 ANK 286.
FT REPEAT 12767 12806 ANK 287.
FT REPEAT 12807 12846 ANK 288.
FT REPEAT 12847 12886 ANK 289.
FT REPEAT 12887 12926 ANK 290.
FT REPEAT 12927 12966 ANK 291.
FT REPEAT 12967 13006 ANK 292.
FT REPEAT 13007 13046 ANK 293.
FT REPEAT 13047 13086 ANK 294.
FT REPEAT 13087 13126 ANK 295.
FT REPEAT 13127 13166 ANK 296.
FT REPEAT 13167 13206 ANK 297.
FT REPEAT 13207 13246 ANK 298.
FT REPEAT 13247 13286 ANK 299.
FT REPEAT 13287 13326 ANK 300.
FT REPEAT 13327 13366 ANK 301.
FT REPEAT 13367 13406 ANK 302.
FT REPEAT 13407 13446 ANK 303.
FT REPEAT 13447 13486 ANK 304.
FT REPEAT 13487 13526 ANK 305.
FT REPEAT 13527 13566 ANK 306.
FT REPEAT 13567 13606 ANK 307.
FT REPEAT 13607 13646 ANK 308.
FT REPEAT 13647 13686 ANK 309.
FT REPEAT 13687 13726 ANK 310.
FT REPEAT 13727 13766 ANK 311.
FT REPEAT 13767 13806 ANK 312.
FT REPEAT 13807 13846 ANK 313.
FT REPEAT 13847 13886 ANK 314.
FT REPEAT 13887 13926 ANK 315.
FT REPEAT 13927 13966 ANK 316.
FT REPEAT 13967 14006 ANK 317.
FT REPEAT 14007 14046 ANK 318.
FT REPEAT 14047 14086 ANK 319.
FT REPEAT 14087 14126 ANK 320.
FT REPEAT 14127 14166 ANK 321.
FT REPEAT 14167 14206 ANK 322.
FT REPEAT 14207 14246 ANK 323.
FT REPEAT 14247 14286 ANK 324.
FT REPEAT 14287 14326 ANK 325.
FT REPEAT 14327 14366 ANK 326.
FT REPEAT 14367 14406 ANK 327.
FT REPEAT 14407 14446 ANK 328.
FT REPEAT 14447 14486 ANK 329.
FT REPEAT 14487 14526 ANK 330.
FT REPEAT 14527 14566 ANK 331.
FT REPEAT 14567 14606 ANK 332.
FT REPEAT 14607 14646 ANK 333.
FT REPEAT 14647 14686 ANK 334.
FT REPEAT 14687 14726 ANK 335.
FT REPEAT 14727 14766 ANK 336.
FT REPEAT 14767 14806 ANK 337.
FT REPEAT 14807 14846 ANK 338.
FT REPEAT 14847 14886 ANK 339.
FT REPEAT 14887 14926 ANK 340.
FT REPEAT 14927 14966 ANK 341.
FT REPEAT 14967 15006 ANK 342.
FT REPEAT 15007 15046 ANK 343.
FT REPEAT 15047 15086 ANK 344.
FT REPEAT 15087 15126 ANK 345.
FT REPEAT 15127 15166 ANK 346.
FT REPEAT 15167 15206 ANK 347.
FT REPEAT 15207 15246 ANK 348.
FT REPEAT 15247 15286 ANK 349.
FT REPEAT 15287 15326 ANK 350.
FT REPEAT 15327 15366 ANK 351.
FT REPEAT 15367 15406 ANK 352.
FT REPEAT 15407 15446 ANK 353.
FT REPEAT 15447 15486 ANK 354.
FT REPEAT 15487 15526 ANK 355.
FT REPEAT 15527 15566 ANK 356.
FT REPEAT 15567 15606 ANK 357.
FT REPEAT 15607 15646 ANK 358.
FT REPEAT 15647 15686 ANK 359.
FT REPEAT 15687 15726 ANK 360.
FT REPEAT 15727 15766 ANK 361.
FT REPEAT 15767 15806 ANK 362.
FT REPEAT 15807 15846 ANK 363.
FT REPEAT 15847 15886 ANK 364.
FT REPEAT 15887 15926 ANK 365.
FT REPEAT 15927 15966 ANK 366.
FT REPEAT 15967 16006 ANK 367.
FT REPEAT 16007 16046 ANK 368.
FT REPEAT 16047 16086 ANK 369.
FT REPEAT 16087 16126 ANK 370.
FT REPEAT 16127 16166 ANK 371.
FT REPEAT 16167 16206 ANK 372.
FT REPEAT 16207 16246 ANK 373.
FT REPEAT 16247 16286 ANK 374.
FT REPEAT 16287 16326 ANK 375.
FT REPEAT 16327 16366 ANK 376.
FT REPEAT 16367 16406 ANK 377.
FT REPEAT 16407 16446 ANK 378.
FT REPEAT 16447 16486 ANK 379.
FT REPEAT 16487 16526 ANK 380.
FT REPEAT 16527 16566 ANK 381.
FT REPEAT 16567 16606 ANK 382.
FT REPEAT 16607 16646 ANK 383.
FT REPEAT 16647 16686 ANK 384.
FT REPEAT 16687 16726 ANK 385.
FT REPEAT 16727 16766 ANK 386.
FT REPEAT 16767 16806 ANK 387.
FT REPEAT 16807 16846 ANK 388.
FT REPEAT 16847 16886 ANK 389.
FT REPEAT 16887 16926 ANK 390.
FT REPEAT 16927 16966 ANK 391.
FT REPEAT 16967 17006 ANK 392.
FT REPEAT 17007 17046 ANK 393.
FT REPEAT 17047 17086 ANK 394.
FT REPEAT 17087 17126 ANK 395.
FT REPEAT 17127 17166 ANK 396.
FT REPEAT 17167 17206 ANK 397.
FT REPEAT 17207 17246 ANK 398.
FT REPEAT 17247 17286 ANK 399.
FT REPEAT 17287 17326 ANK 400.
FT REPEAT 17327 17366 ANK 401.
FT REPEAT 17367 17406 ANK 402.
FT REPEAT 17407 17446 ANK 403.
FT REPEAT 17447 17486 ANK 404.
FT REPEAT 17487 17526 ANK 405.
FT REPEAT 17527 17566 ANK 406.
FT REPEAT 17567 17606 ANK 407.
FT REPEAT 17607 17646 ANK 408.
FT REPEAT 17647 17686 ANK 409.
FT REPEAT 17687 17726 ANK 410.
FT REPEAT 17727 17766 ANK 411.
FT REPEAT 17767 17806 ANK 412.
FT REPEAT 17807 17846 ANK 413.
FT REPEAT 17847 17886 ANK 414.
FT REPEAT 17887 17926 ANK 415.
FT REPEAT 17927 17966 ANK 416.
FT REPEAT 17967 18006 ANK 417.
FT REPEAT 18007 18046 ANK 418.
FT REPEAT 18047 18086 ANK 419.
FT REPEAT 18087 18126 ANK 420.
FT REPEAT 18127 18166 ANK 421.
FT REPEAT 18167 18206 ANK 422.
FT REPEAT 18207 18246 ANK 423.
FT REPEAT 18247 18286 ANK 424.
FT REPEAT 18287 18326 ANK 425.
FT REPEAT 18327 18366 ANK 426.
FT REPEAT 18367 18406 ANK 427.
FT REPEAT 18407 18446 ANK 428.
FT REPEAT 18447 18486 ANK 429.
FT REPEAT 18487 18526 ANK 430.
FT REPEAT 18527 18566 ANK 431.
FT REPEAT 18567 18606 ANK 432.
FT REPEAT 18607 18646 ANK 433.
FT REPEAT 18647 18686 ANK 434.
FT REPEAT 18687 18726 ANK 435.
FT REPEAT 18727 18766 ANK 436.
FT REPEAT 18767 18806 ANK 437.
FT REPEAT 18807 18846 ANK 438.
FT REPEAT 18847 18886 ANK 439.
FT REPEAT 18887 18926 ANK 440.
FT REPEAT 18927 18966 ANK 441.
FT REPEAT 18967 19006 ANK 442.
FT REPEAT 19007 19046 ANK 443.
FT REPEAT 19047 19086 ANK 444.
FT REPEAT 19087 19126 ANK 445.
FT REPEAT 19127 19166 ANK 446.
FT REPEAT 19167 19206 ANK 447.
FT REPEAT 19207 19246 ANK 448.
FT REPEAT 19247 19286 ANK 449.
FT REPEAT 19287 19326 ANK 450.
FT REPEAT 19327 19366 ANK 451.
FT REPEAT 19367 19406 ANK 452.
FT REPEAT 19407 19446 ANK 453.
FT REPEAT 19447 19486 ANK 454.
FT REPEAT 19487 19526 ANK 455.
FT REPEAT 19527 19566 ANK 456.
FT REPEAT 19567 19606 ANK 457.
FT REPEAT 19607 19646 ANK 458.
FT REPEAT 19647 19686 ANK 459.
FT REPEAT 19687 19726 ANK 460.
FT REPEAT 19727 19766 ANK 461.
FT REPEAT 19767 19806 ANK 462.
FT REPEAT 19807 19846 ANK 463.
FT REPEAT 19847 19886 ANK 464.
FT REPEAT 19887 19926 ANK 465.
FT REPEAT 19927 19966 ANK 466.
FT REPEAT 19967 20006 ANK 467.
FT REPEAT 20007 20046 ANK 468.
FT REPEAT 20047 20086 ANK 469.
FT REPEAT 20087 20126 ANK 470.
FT REPEAT 20127 20166 ANK 471.
FT REPEAT 20167 20206 ANK 472.
FT REPEAT 20207 20246 ANK 473.
FT REPEAT 20247 20286 ANK 474.
FT REPEAT 20287 20326 ANK 475.
FT REPEAT 20327 20366 ANK 476.
FT REPEAT 20367 20406 ANK 477.
FT REPEAT 20407 20446 ANK 478.
FT REPEAT 20447 20486 ANK 479.
FT REPEAT 20487 20526 ANK 480.
FT REPEAT 20527 20566 ANK 481.
FT REPEAT 20567 20606 ANK 482.
FT REPEAT 20607 20646 ANK 483.
FT REPEAT 20647 20686 ANK 484.
FT REPEAT 20687 20726 ANK 485.
FT REPEAT 20727 20766 ANK 486.
FT REPEAT 20767 20806 ANK 487.
FT REPEAT 20807 20846 ANK 488.
FT REPEAT 20847 20886 ANK 489.
FT REPEAT 20887 20926 ANK 490.
FT REPEAT 20927 20966 ANK 491.
FT REPEAT 20967 21006 ANK 492.
FT REPEAT 21007 21046 ANK 493.
FT REPEAT 21047 21086 ANK 494.
FT REPEAT 21087 21126 ANK 495.
FT REPEAT 21127 21166 ANK 496.
FT REPEAT 21167 21206 ANK 497.
FT REPEAT 21207 21246 ANK 498.
FT REPEAT 21247 21286 ANK 499.
FT REPEAT 21287 21326 ANK 500.
FT REPEAT 21327 21366 ANK 501.
FT REPEAT 21367 21406 ANK 502.
FT REPEAT 21407 21446 ANK 503.
FT REPEAT 21447 21486 ANK 504.
FT REPEAT 21487 21526 ANK 505.
FT REPEAT 21527 21566 ANK 506.
FT REPEAT 21567 21606 ANK 507.
FT REPEAT 21607 21646 ANK 508.
FT REPEAT 21647 21686 ANK 509.
FT REPEAT 21687 21726 ANK 510.
FT REPEAT 21727 21766 ANK 511.
FT REPEAT 21767 21806 ANK 512.
FT REPEAT 21807 21846 ANK 513.
FT REPEAT 21847 21886 ANK 514.
FT REPEAT 21887 21926 ANK 515.
FT REPEAT 21927 21966 ANK 516.
FT REPEAT 21967 22006 ANK 517.
FT REPEAT 22007 22046 ANK 518.
FT REPEAT 22047 22086 ANK 519.
FT REPEAT 22087 22126 ANK 520.
FT REPEAT 22127 22166 ANK 521.
FT REPEAT 22167 22206 ANK 522.
FT REPEAT 22207 22246 ANK 523.
FT REPEAT 22247 22286 ANK 524.
FT REPEAT 22287 22326 ANK 525.
FT REPEAT 22327 22366 ANK 526.
FT REPEAT 22367 22406 ANK 527.
FT REPEAT 22407 22446 ANK 528.
FT REPEAT 22447 22486 ANK 529.
FT REPEAT 22487 22526 ANK 530.
FT REPEAT 22527 22566 ANK 531.
FT REPEAT 22567 22606 ANK 532.
FT REPEAT 22607 22646 ANK 533.
FT REPEAT 22647 22686 ANK 534.
FT REPEAT 22687 22726 ANK 535.
FT REPEAT 22727 22766 ANK 536.
FT REPEAT 22767 22806 ANK 537.
FT REPEAT 22807 22846 ANK 538.
FT REPEAT 22847 22886 ANK 539.
FT REPEAT 22887 22926 ANK 540.
FT REPEAT 22927 22966 ANK 541.
FT REPEAT 22967 23006 ANK 542.
FT REPEAT 23007 23046 ANK 543.
FT REPEAT 23047 23086 ANK 544.
FT REPEAT 23087 23126 ANK 545.
FT REPEAT 23127 23166 ANK 546.
FT REPEAT
```

FT	DOMAIN	1020	1056	EGF-LIKE 27,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1058	1094	EGF-LIKE 28,	
FT	DOMAIN	1096	1142	EGF-LIKE 29,	
FT	DOMAIN	1144	1180	EGF-LIKE 30,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1182	1218	EGF-LIKE 31,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1220	1264	EGF-LIKE 32,	
FT	DOMAIN	1266	1304	EGF-LIKE 33,	
FT	DOMAIN	1306	1346	EGF-LIKE 34,	
FT	DOMAIN	1347	1383	EGF-LIKE 35,	
FT	DOMAIN	1386	1424	EGF-LIKE 36,	
FT	REPEAT	1441	1478	LIN/NOTCH 1,	
FT	REPEAT	1479	1520	LIN/NOTCH 2,	
FT	REPEAT	1521	1560	LIN/NOTCH 3,	
FT	REPEAT	1876	1919	ANK 1,	
FT	REPEAT	1924	1953	ANK 2,	
FT	REPEAT	1957	1987	ANK 3,	
FT	REPEAT	1991	2020	ANK 4,	
FT	REPEAT	2024	2053	ANK 5,	
FT	REPEAT	2057	2086	ANK 6,	
FT	DISULFID	22	35	BY SIMILARITY,	
FT	DISULFID	29	45	BY SIMILARITY,	
FT	DISULFID	47	56	BY SIMILARITY,	
FT	DISULFID	62	74	BY SIMILARITY,	
FT	DISULFID	68	87	BY SIMILARITY,	
FT	DISULFID	89	98	BY SIMILARITY,	
FT	DISULFID	106	117	BY SIMILARITY,	
FT	DISULFID	111	128	BY SIMILARITY,	
FT	DISULFID	130	139	BY SIMILARITY,	
FT	DISULFID	145	156	BY SIMILARITY,	
FT	DISULFID	150	165	BY SIMILARITY,	
FT	DISULFID	167	176	BY SIMILARITY,	
FT	DISULFID	183	194	BY SIMILARITY,	
FT	DISULFID	188	203	BY SIMILARITY,	
FT	DISULFID	205	214	BY SIMILARITY,	
FT	DISULFID	221	232	BY SIMILARITY,	
FT	DISULFID	226	242	BY SIMILARITY,	
FT	DISULFID	244	253	BY SIMILARITY,	
FT	DISULFID	260	271	BY SIMILARITY,	
FT	DISULFID	265	280	BY SIMILARITY,	
FT	DISULFID	282	291	BY SIMILARITY,	
FT	DISULFID	298	311	BY SIMILARITY,	
FT	DISULFID	305	320	BY SIMILARITY,	
FT	DISULFID	322	331	BY SIMILARITY,	
FT	DISULFID	338	349	BY SIMILARITY,	
FT	DISULFID	343	358	BY SIMILARITY,	
FT	DISULFID	360	369	BY SIMILARITY,	
FT	DISULFID	375	386	BY SIMILARITY,	
FT	DISULFID	380	397	BY SIMILARITY,	
FT	DISULFID	399	408	BY SIMILARITY,	
FT	DISULFID	415	428	BY SIMILARITY,	
FT	DISULFID	422	437	BY SIMILARITY,	
FT	DISULFID	439	448	BY SIMILARITY,	
FT	DISULFID	455	466	BY SIMILARITY,	
FT	DISULFID	460	475	BY SIMILARITY,	
FT	DISULFID	477	486	BY SIMILARITY,	
FT	DISULFID	493	504	BY SIMILARITY,	
FT	DISULFID	498	513	BY SIMILARITY,	
FT	DISULFID	515	524	BY SIMILARITY,	
FT	DISULFID	531	542	BY SIMILARITY,	
FT	DISULFID	536	551	BY SIMILARITY,	
FT	DISULFID	553	562	BY SIMILARITY,	
FT	DISULFID	569	579	BY SIMILARITY,	
FT	DISULFID	574	588	BY SIMILARITY,	
FT	DISULFID	590	599	BY SIMILARITY,	
FT	DISULFID	606	617	BY SIMILARITY,	
FT	DISULFID	611	626	BY SIMILARITY,	
FT	DISULFID	628	637	BY SIMILARITY,	
FT	DISULFID	644	654	BY SIMILARITY,	
FT	DISULFID	649	663	BY SIMILARITY,	
FT	DISULFID	665	674	BY SIMILARITY,	
FT	DISULFID	681	692	BY SIMILARITY,	
FT	DISULFID	686	701	BY SIMILARITY,	
FT	DISULFID	703	712	BY SIMILARITY,	

FT	DISULFID	719	729	BY SIMILARITY,	
FT	DISULFID	724	738	BY SIMILARITY,	
FT	DISULFID	740	748	BY SIMILARITY,	
FT	DISULFID	756	767	BY SIMILARITY,	
FT	DISULFID	761	776	BY SIMILARITY,	
FT	DISULFID	778	787	BY SIMILARITY,	
FT	DISULFID	794	805	BY SIMILARITY,	
FT	DISULFID	799	814	BY SIMILARITY,	
FT	DISULFID	816	825	BY SIMILARITY,	
FT	DISULFID	832	843	BY SIMILARITY,	
FT	DISULFID	837	854	BY SIMILARITY,	
FT	DISULFID	856	865	BY SIMILARITY,	
FT	DISULFID	872	883	BY SIMILARITY,	
FT	DISULFID	877	892	BY SIMILARITY,	
FT	DISULFID	894	903	BY SIMILARITY,	
FT	DISULFID	910	921	BY SIMILARITY,	
FT	DISULFID	915	930	BY SIMILARITY,	
FT	DISULFID	932	941	BY SIMILARITY,	
FT	DISULFID	966	997	BY SIMILARITY,	
FT	DISULFID	991	1006	BY SIMILARITY,	
FT	DISULFID	1008	1017	BY SIMILARITY,	
FT	DISULFID	1024	1035	BY SIMILARITY,	
FT	DISULFID	1029	1044	BY SIMILARITY,	
FT	DISULFID	1046	1055	BY SIMILARITY,	
FT	DISULFID	1062	1073	BY SIMILARITY,	
FT	DISULFID	1067	1082	BY SIMILARITY,	
FT	DISULFID	1084	1093	BY SIMILARITY,	
FT	DISULFID	1100	1121	BY SIMILARITY,	
FT	DISULFID	1115	1130	BY SIMILARITY,	

Query Match 3.0%; Score 149.5; DB 1; Length 2524;
 Best Local Similarity 19.6%; Pred. No. 0.67; Matches 143; Conservative 95; Mismatches 217; Indels 275; Gaps 42;

Qy	412	GRDWPDPFSSKQKVVLTWVPNCABGCGSGWIKDG-YDCKAUNNSACDMDGDCSG	470
Db	1435	GKNTTPDDNDICENQ-----CSE-----LADKVCNANCNNAACGMDGDC--	1477
Qy	471	NSGSRVYAGGGTSGVGHFWQ-----FGGINSVSVCN-GCA-----	510
Db	1478	-----SLNFNDPWNKCTGSLQCKRYFNDD-KDSCQNNNGCYDGFDCQ	1521
Qy	511	-----NSWLADKF-----CDQACNVASCDFDAGDCQ--DHPIE-LYVYILP--	550
Db	1522	VEVQCNPVLDQYKDFHGDGCHDQCGNNAECMDGLDCAENMENAEGTLVIVLMPER	1581
Qy	551	---NQHVI-----IPKGE-CLPYF-SFAEVAKGVGGA---YSDNP--	584
Db	1582	LKNNSVNFRELSRVLTWVPFKOSKGEYKIPYVYGNNEELKHKIKRSTWSDAPSA	1641
Qy	585	-----IIRHASTANK--WKTILHMSGMNATTIHNLTF	617
Db	1642	ISTKTESILLGHRRELDMEVRSIVYLEIDNRQCYSSQCSNATDVAFGLALAS	1701
Qy	618	QNTNDE-EFMQITVEVDTRREGPKLNSTAKGVENLVSEITLL-----DEA	662
Db	1702	LGSLDLTSLYKIE-AVXSENNMETPK--PSTLYPMLMLVPLIILFVFMVIVNKKRRREG	1759
Qy	663	EILFED--IPKGRFKFKGRHD-----	686
Db	1760	QLMFPDGFIPKEP--SKKRRDRRLGEDSVGLKPIKNMTDGSFMDNQNQEWGDEETLENKR	1817
Qy	687	RRAQSEVKIP-LVNISILPKD-----AQLSLNTL-----DLQLEHGDTITKG--	727
Db	1818	FRFEEGVILLPELVDDTDPRQMTROHLDAADLRISMATPPOGRIEADCMVNVRRGPDG	1877
Qy	728	-----YLSKSAILRFSLNNSQAKIKNOAIIITDENDSLVAQ	766
Db	1878	FTPLMLASCSGGGLGTGNSSEEDASANNISDPI--GQGAQLANOTDRGTALHLAAAY	1935
Qy	767	EK-QVHKSILPNS--LGVSERLQRLTFPPAVSVKNGHDG-----QNPPLDET-----	812
Db	1936	ARADAKRLBESSADANVQDNMGRTPLHA---VAADAGVYQIILIRARATLDDARMPDG	1992

```
QY 813 -----TARFVETHTQKTTGG-----NTKKEKPS-----LIYPLBSQMTKEKKTQKE 856
Dy 1993 TTPILARLAVEGVNEELINAHADVNAVDEFGKSAHMAAVNNVDAVLLKNSANKD 2052
QY 857 KENSHMEEN-----AENHIGTEVLLGRKLOHYTD-----SYLGLPWE--KKKTFOD--- 902
Dy 2053 MNNNEETSLFLAANESEYETAKVL---LDHYANRDTIDMDLPPROIAGERHHIIVH 2108
QY 903 LIDEEESLKT 912
Dy 2109 LIDEXNLVKS 2118

RESULT 12
NTG3_RAT
ID NTG3_RAT STANDARD; PRT; 2319 AA.
AC 09R172;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN NOTCH3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann T., Boulter J., Weinmaster G., Schanen N.C.;
RT "Rattus norvegicus mRNA for Notch 3."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP FUNCTION.
RA MEDLINE=21094508; PubMed=11182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA Honjo T.;
RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT neural progenitor cells to an astroglial fate."
RL Neuron 29:45-55(2001).
[3]
RP TISSUE SPECIFICITY.
RX MEDLINE=21331789; PubMed=11438922;
RX Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT functional roles for the Notch-Delta signaling system during brain
RT development."
RL J. Comp. Neurol. 436:167-181(2001).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). Acts instructively to control
CC the cell fate determination of CNS multipotent progenitor cells,
CC resulting in astroglial induction and neuron/oligodendrocyte
CC suppression.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- TISSUE SPECIFICITY: Expressed in postnatal central nervous system
CC (CNS) germinal zones and, in early postnatal life, within
CC numerous cells throughout the CNS. It is more highly localized
CC to ventricular germinal zones.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like cleavage in the
CC trans-golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
```

FT	DOMAIN	584	620	EGF-LIKE 15,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	622	657	EGF-LIKE 16,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	659	695	EGF-LIKE 17,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	697	732	EGF-LIKE 18,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	736	772	EGF-LIKE 19,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	773	810	EGF-LIKE 20,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	812	849	EGF-LIKE 21,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	851	887	EGF-LIKE 22,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	889	924	EGF-LIKE 23,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	962	962	EGF-LIKE 24,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	964	1000	EGF-LIKE 25,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1002	1036	EGF-LIKE 26,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1038	1084	EGF-LIKE 27,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1086	1122	EGF-LIKE 28,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1124	1160	EGF-LIKE 29,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1162	1205	EGF-LIKE 30,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1207	1246	EGF-LIKE 31,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1248	1289	EGF-LIKE 32,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1291	1327	EGF-LIKE 33,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1337	1375	EGF-LIKE 34,	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1384	1420	LIN/NOTCH 1,	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1425	1461	LIN/NOTCH 2,	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1469	1503	LIN/NOTCH 3,	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1540	1569	ANK 1,	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1573	1593	ANK 2,	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1607	1636	ANK 3,	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1699	1733	ANK 4,	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1973	2002	ANK 5,	CALCIUM-BINDING (POTENTIAL)
FT	SITE	1573	1574	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY),	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	44	56	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	50	67	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	69	78	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	84	95	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	89	108	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	110	119	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	125	136	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	130	146	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	148	157	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	164	176	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	170	185	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	187	196	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	203	214	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	208	224	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	226	235	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	242	253	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	247	262	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	264	273	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	280	293	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	287	302	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	304	313	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	320	331	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	325	340	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	342	351	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	357	368	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	362	379	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	381	390	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	397	410	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	404	419	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	421	430	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	437	448	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	442	457	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	459	468	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	475	486	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	480	495	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	497	506	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	513	524	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	518	533	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	535	544	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	551	561	BY SIMILARITY,	CALCIUM-BINDING (POTENTIAL)

Query Match 3.0%
Best Local Similarity 25.9%
Score 148; DB 1; Length 2319;
Pred. No. 0.73;

Matches	36;	Conservative	12;	Mismatches	35;	Indels	56;	Gaps	5;
Qy	436	PNCABGCPGSMITDGYCDACNNACDPMGHCNSGSGSRYIAGGCTGSGVGHFPO-	494						
Db	1387	PRCPRAACQAKRGDQDCRECNPSGCGMGCGDC-----SINVDPVWQ	1429						
Qy	495	-----GGGINSYSYCN-----QGCANSLWADKPC	519						
Db	1430	CEALQCMRLFNNSRCDPACSSPACLYDNDPCISGGDRP--GNPYKTKCALHF-ADGRG	1486						
Qy	520	DQACNVLSGCFPDAGCGGD	538						
Db	1487	DQGCNTEGCGWGLDCASE	1505						
RESULT 13									
NTC3_MOUSE									
ID	NTC3_MOUSE	STANDARD;	PRT;	2318	AA.				
AC	Q61982;								
DT	01-NOV-1997 (Rel. 35, Created)								
DT	01-NOV-1997 (Rel. 35, Last sequence update)								
DT	28-FEB-2003 (Rel. 41, Last annotation update)								
DE	Neurogenic locus notch homolog protein 3 precursor (Notch 3).								
GN	NOTCH3.								
OS	Mus musculus (Mouse).								
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
OX	NCBI_TaxId=10090;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=ICR X Swiss Webster;								
RX	MEDLINE=95001556; PubMed=7918097;								
RA	Lardelli M., Balstrand J., Lendahl U.;								
RT	"The novel Notch homolog mouse Notch 3 lacks specific epidermal								
RT	growth factor-repeats and is expressed in proliferating								
RT	neuroepithelium".								
RL	Mech. Dev. 46:123-136(1994).								
RN	[2]								
RP	POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1664.								
RX	MEDLINE=21523956; PubMed=11518718;								
RA	Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;								
RT	"Murine notch homologs (NL-4) undergo presenilin-dependent								
RT	proteolysis".								
RL	J. Biol. Chem. 276:40268-40273(2001).								
RN	[3]								
RP	POST-TRANSLATIONAL PROCESSING.								
RX	MEDLINE=21374376; PubMed=11459941;								
RA	Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;								
RT	"Conservation of the biochemical mechanisms of signal transduction								
RT	among mammalian Notch family members".								
RL	Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).								
CC	-1- FUNCTION: Functions as a receptor for membrane-bound ligands								
CC	Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.								
CC	Upon ligand activation through the released notch intracellular								
CC	domain (NICD) it forms a transcriptional activator complex with								
CC	RBP-J kappa and activates genes of the enhancer of split locus.								
CC	Affects the implementation of differentiation, proliferation and								
CC	apoptotic programs (By similarity). May play a role during CNS								
CC	development.								
CC	-1- SUBUNIT: Heterodimer of a C-terminal fragment N(Tm) and a N-								
CC	terminal fragment N(IEC) which are probably linked by disulfide								
CC	bonds.								
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein. Following								
CC	proteolytical processing NICD is translocated to the nucleus.								
CC	-1- TISSUE SPECIFICITY: Proliferating neuroepithelium.								
CC	-1- DEVELOPMENTAL STAGE: CNS development.								
CC	-1- PTM: Synthesized in the endoplasmic reticulum as an inactive form								
CC	which is proteolytically cleaved by a furin-like convertase in the								
CC	trans-golgi network before it reaches the plasma membrane to yield								
CC	an active, ligand-accessible form. Cleavage results in a C-								
CC	terminal fragment N(Tm) and a N-terminal fragment N(IEC). Following								
CC	ligand binding, it is cleaved by TNF-alpha converting enzyme								
CC	(TACE) to yield a membrane-associated intermediate fragment called								

Best Local Similarity 25.9%; Pred. No. 0.85;
Matches 36; Conservative 11; Mismatches 36; Indels 56; Gaps 5;

QY 436 PNCAGCGPGSWIKDGYCDKACNNNSACDMDGDCSGNSGSRRIAGGSGTSGVCHPQC- 494
1386 PRCPRACQAKXRGDQNCRECTPGCGMDGDC-----SUNVDDPRQ 1428

QY 495 -----FGGINSVSYCN-----GCGANSLADKFC 519
1429 CEALQCRLENNNSRCDPACSPACLYDNPCYSGGRDRT--CNFYKXCADHF-ADGRC 1485

QY 520 DQACNVLSCGPDAGCGCD 538
1486 DQGCNTEBCGMDGLDCASE 1504

Db

RESULT 14
NTCL HUMAN
ID NTCL HUMAN STANDARD; PRT: 2556 AA.

AC P46531;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (hml)
DE (Translocation-associated notch protein TAN-1).
GN NOTCH1 OR TAN1
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Mann R.S., Blummeier C.M., Zagouras P.;
RT "Complete human notch 1 (hml) cDNA sequence."
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE OF 1-2444 FROM N.A.
RA MEDLINE=91347367; PubMed=1831692;
RX Eilisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RT Smith S.D., Sklar J.;
RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT chromosomal translocations in T lymphoblastic neoplasms."
RL Cell 66:649-661(1991).
RN (3)
RP IDENTIFICATION OF LIGANDS.
RA MEDLINE=99180765; PubMed=10079256;
RX Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.L.,
RA Banks A., Leiman U., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT "Human ligands of the Notch receptor."
RL Am. J. Pathol. 154:785-794(1999).
RN (4)
RP INTERACTION WITH DTX1.
RX MEDLINE=98250176; PubMed=9590294;
RA Matsuno K., Eastman D., Mitsiadis T., Quinn A.M., Carcangiu M.L.,
RA Ordenitch P., Kadesch T., Artavanis-Tsakonas S.;
RT "Human deltex is a conserved regulator of Notch signaling."
RL Nat. Genet. 19:74-78(1998).
RN (5)
RP FUNCTION: Functions as a receptor for membrane-bound ligands
Jagged2, Jagged2 and Delta1 to regulate cell-fate determination.
Upon ligand activation through the released notch intracellular
domain (NICD) it forms a transcriptional activator complex with
RBP-J/Kap1 and activates genes of the enhancer of split locus.
Affects the implementation of differentiation, proliferation and
apoptotic programs. May be important for normal lymphocyte
function. In altered form, may contribute to transformation or
progression in some T-cell neoplasms. Involved in the maturation
of both CD4+ and CD8+ cells in the thymus (By similarity).
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
terminal fragment N(IEC) which are probably linked by disulfide
bonds (By similarity). Interacts with DTX1 and DTX2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
proteolytic processing NICD is translocated to the nucleus (By

CC similarity).
CC -1- TISSUE SPECIFICITY: In fetal tissues most abundant in spleen,
CC brain stem and lung. Also present in most adult tissues where it
CC is found mainly in lymphoid tissues.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(IEC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXTR). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -1- PTM: Phosphorylated (By similarity).
CC -1- DISEASE: NOTCH1 truncation is associated with T-cell acute
CC lymphoblastic leukemia.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 36 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 5 ANK repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF308602; AA033848.1; -.
DR EMBL: M73980; AA06014.1; -.
DR HSSP: P00740; IEDM.
DR Genew: HGNC:7881; NOTCH1.
DR MIM: 190198; -.
DR GO: GO:0016021; C:integral to membrane; NAS.
DR GO: GO:0003793; P:defense/immunity protein activity; NAS.
DR GO: GO:000655; P:immune response; NAS.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl_S.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR002049; Lamln_EGF.
DR InterPro: IPR008287; Notch.
DR InterPro: IPR008000; Notch_dom.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00066; notch; 3.
DR PIRSF: PIRSF002279; Notch; 1.
DR PRINTS: PR00010; EGFLOOD.
DR PRINTS: PR00011; EGFAMININ.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 6.
DR SMART: SM00179; EGF_CA; 23.
DR SMART: SM00004; NL; 3.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS50088; ANK_REPEAT; 4.
DR PROSITE: PS50010; ASK_HYDROXYL; 20.
DR PROSITE: PS00022; EGF_1; 34.
DR PROSITE: PS01186; EGF_2; 26.
DR PROSITE: PS50026; EGF_3; 36.
DR PROSITE: PS01187; EGF_CA; 18.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 18
FT CHAIN 19 2556 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT CHAIN 1722 2556 NOTCH EXTRACELLULAR TRUNCATION (BY
FT SIMILARITY).
FT CHAIN 1755 2556 NOTCH INTRACELLULAR DOMAIN (BY

FT	DOMAIN	19	1736	SIMILARITY).
FT	TRASMEN	1737	1757	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	1758	2556	POTENTIAL.
FT	DOMAIN	20	58	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	59	99	EGF-LIKE 1.
FT	DOMAIN	102	139	EGF-LIKE 2.
FT	DOMAIN	140	176	EGF-LIKE 3.
FT	DOMAIN	148	176	EGF-LIKE 4.
FT	DOMAIN	218	255	EGF-LIKE 5.
FT	DOMAIN	257	293	EGF-LIKE 6.
FT	DOMAIN	293	333	EGF-LIKE 7.
FT	DOMAIN	333	371	EGF-LIKE 8.
FT	DOMAIN	371	410	EGF-LIKE 9.
FT	DOMAIN	410	450	EGF-LIKE 10.
FT	DOMAIN	450	488	EGF-LIKE 11.
FT	DOMAIN	488	526	EGF-LIKE 12.
FT	DOMAIN	526	564	EGF-LIKE 13.
FT	DOMAIN	564	601	EGF-LIKE 14.
FT	DOMAIN	601	639	EGF-LIKE 15.
FT	DOMAIN	639	676	EGF-LIKE 16.
FT	DOMAIN	676	714	EGF-LIKE 17.
FT	DOMAIN	714	751	EGF-LIKE 18.
FT	DOMAIN	751	789	EGF-LIKE 19.
FT	DOMAIN	789	827	EGF-LIKE 20.
FT	DOMAIN	827	868	EGF-LIKE 21.
FT	DOMAIN	868	906	EGF-LIKE 22.
FT	DOMAIN	906	944	EGF-LIKE 23.
FT	DOMAIN	944	982	EGF-LIKE 24.
FT	DOMAIN	982	1020	EGF-LIKE 25.
FT	DOMAIN	1020	1058	EGF-LIKE 26.
FT	DOMAIN	1058	1096	EGF-LIKE 27.
FT	DOMAIN	1096	1134	EGF-LIKE 28.
FT	DOMAIN	1134	1172	EGF-LIKE 29.
FT	DOMAIN	1172	1210	EGF-LIKE 30.
FT	DOMAIN	1210	1248	EGF-LIKE 31.
FT	DOMAIN	1248	1286	EGF-LIKE 32.
FT	DOMAIN	1286	1324	EGF-LIKE 33.
FT	DOMAIN	1324	1362	EGF-LIKE 34.
FT	DOMAIN	1362	1400	EGF-LIKE 35.
FT	DOMAIN	1400	1438	EGF-LIKE 36.
FT	DOMAIN	1438	1476	EGF-LIKE 37.
FT	DOMAIN	1476	1514	EGF-LIKE 38.
FT	DOMAIN	1514	1552	EGF-LIKE 39.
FT	DOMAIN	1552	1590	EGF-LIKE 40.
FT	DOMAIN	1590	1628	EGF-LIKE 41.
FT	DOMAIN	1628	1666	EGF-LIKE 42.
FT	DOMAIN	1666	1704	EGF-LIKE 43.
FT	DOMAIN	1704	1742	EGF-LIKE 44.
FT	DOMAIN	1742	1780	EGF-LIKE 45.
FT	DOMAIN	1780	1818	EGF-LIKE 46.
FT	DOMAIN	1818	1856	EGF-LIKE 47.
FT	DOMAIN	1856	1894	EGF-LIKE 48.
FT	DOMAIN	1894	1932	EGF-LIKE 49.
FT	DOMAIN	1932	1970	EGF-LIKE 50.
FT	DOMAIN	1970	2008	EGF-LIKE 51.
FT	DOMAIN	2008	2046	EGF-LIKE 52.
FT	DOMAIN	2046	2084	EGF-LIKE 53.
FT	DOMAIN	2084	2122	EGF-LIKE 54.
FT	DOMAIN	2122	2160	EGF-LIKE 55.
FT	DOMAIN	2160	2198	EGF-LIKE 56.
FT	DOMAIN	2198	2236	EGF-LIKE 57.
FT	DOMAIN	2236	2274	EGF-LIKE 58.
FT	DOMAIN	2274	2312	EGF-LIKE 59.
FT	DOMAIN	2312	2350	EGF-LIKE 60.
FT	DOMAIN	2350	2388	EGF-LIKE 61.
FT	DOMAIN	2388	2426	EGF-LIKE 62.
FT	DOMAIN	2426	2464	EGF-LIKE 63.
FT	DOMAIN	2464	2502	EGF-LIKE 64.
FT	DOMAIN	2502	2540	EGF-LIKE 65.
FT	DOMAIN	2540	2578	EGF-LIKE 66.
FT	DOMAIN	2578	2616	EGF-LIKE 67.
FT	DOMAIN	2616	2654	EGF-LIKE 68.
FT	DOMAIN	2654	2692	EGF-LIKE 69.
FT	DOMAIN	2692	2730	EGF-LIKE 70.
FT	DOMAIN	2730	2768	EGF-LIKE 71.
FT	DOMAIN	2768	2806	EGF-LIKE 72.
FT	DOMAIN	2806	2844	EGF-LIKE 73.
FT	DOMAIN	2844	2882	EGF-LIKE 74.
FT	DOMAIN	2882	2920	EGF-LIKE 75.
FT	DOMAIN	2920	2958	EGF-LIKE 76.
FT	DOMAIN	2958	2996	EGF-LIKE 77.
FT	DOMAIN	2996	3034	EGF-LIKE 78.
FT	DOMAIN	3034	3072	EGF-LIKE 79.
FT	DOMAIN	3072	3110	EGF-LIKE 80.
FT	DOMAIN	3110	3148	EGF-LIKE 81.
FT	DOMAIN	3148	3186	EGF-LIKE 82.
FT	DOMAIN	3186	3224	EGF-LIKE 83.
FT	DOMAIN	3224	3262	EGF-LIKE 84.
FT	DOMAIN	3262	3300	EGF-LIKE 85.
FT	DOMAIN	3300	3338	EGF-LIKE 86.
FT	DOMAIN	3338	3376	EGF-LIKE 87.
FT	DOMAIN	3376	3414	EGF-LIKE 88.
FT	DOMAIN	3414	3452	EGF-LIKE 89.
FT	DOMAIN	3452	3490	EGF-LIKE 90.
FT	DOMAIN	3490	3528	EGF-LIKE 91.
FT	DOMAIN	3528	3566	EGF-LIKE 92.
FT	DOMAIN	3566	3604	EGF-LIKE 93.
FT	DOMAIN	3604	3642	EGF-LIKE 94.
FT	DOMAIN	3642	3680	EGF-LIKE 95.
FT	DOMAIN	3680	3718	EGF-LIKE 96.
FT	DOMAIN	3718	3756	EGF-LIKE 97.
FT	DOMAIN	3756	3794	EGF-LIKE 98.
FT	DOMAIN	3794	3832	EGF-LIKE 99.
FT	DOMAIN	3832	3870	EGF-LIKE 100.
FT	DOMAIN	3870	3908	EGF-LIKE 101.
FT	DOMAIN	3908	3946	EGF-LIKE 102.
FT	DOMAIN	3946	3984	EGF-LIKE 103.
FT	DOMAIN	3984	4022	EGF-LIKE 104.
FT	DOMAIN	4022	4060	EGF-LIKE 105.
FT	DOMAIN	4060	4098	EGF-LIKE 106.
FT	DOMAIN	4098	4136	EGF-LIKE 107.
FT	DOMAIN	4136	4174	EGF-LIKE 108.
FT	DOMAIN	4174	4212	EGF-LIKE 109.
FT	DOMAIN	4212	4250	EGF-LIKE 110.
FT	DOMAIN	4250	4288	EGF-LIKE 111.
FT	DOMAIN	4288	4326	EGF-LIKE 112.
FT	DOMAIN	4326	4364	EGF-LIKE 113.
FT	DOMAIN	4364	4402	EGF-LIKE 114.
FT	DOMAIN	4402	4440	EGF-LIKE 115.
FT	DOMAIN	4440	4478	EGF-LIKE 116.
FT	DOMAIN	4478	4516	EGF-LIKE 117.
FT	DOMAIN	4516	4554	EGF-LIKE 118.
FT	DOMAIN	4554	4592	EGF-LIKE 119.
FT	DOMAIN	4592	4630	EGF-LIKE 120.
FT	DOMAIN	4630	4668	EGF-LIKE 121.
FT	DOMAIN	4668	4706	EGF-LIKE 122.
FT	DOMAIN	4706	4744	EGF-LIKE 123.
FT	DOMAIN	4744	4782	EGF-LIKE 124.
FT	DOMAIN	4782	4820	EGF-LIKE 125.
FT	DOMAIN	4820	4858	EGF-LIKE 126.
FT	DOMAIN	4858	4896	EGF-LIKE 127.
FT	DOMAIN	4896	4934	EGF-LIKE 128.
FT	DOMAIN	4934	4972	EGF-LIKE 129.
FT	DOMAIN	4972	5010	EGF-LIKE 130.
FT	DOMAIN	5010	5048	EGF-LIKE 131.
FT	DOMAIN	5048	5086	EGF-LIKE 132.
FT	DOMAIN	5086	5124	EGF-LIKE 133.
FT	DOMAIN	5124	5162	EGF-LIKE 134.
FT	DOMAIN	5162	5200	EGF-LIKE 135.
FT	DOMAIN	5200	5238	EGF-LIKE 136.
FT	DOMAIN	5238	5276	EGF-LIKE 137.
FT	DOMAIN	5276	5314	EGF-LIKE 138.
FT	DOMAIN	5314	5352	EGF-LIKE 139.
FT	DOMAIN	5352	5390	EGF-LIKE 140.
FT	DOMAIN	5390	5428	EGF-LIKE 141.
FT	DOMAIN	5428	5466	EGF-LIKE 142.
FT	DOMAIN	5466	5504	EGF-LIKE 143.
FT	DOMAIN	5504	5542	EGF-LIKE 144.
FT	DOMAIN	5542	5580	EGF-LIKE 145.
FT	DOMAIN	5580	5618	EGF-LIKE 146.
FT	DOMAIN	5618	5656	EGF-LIKE 147.
FT	DOMAIN	5656	5694	EGF-LIKE 148.
FT	DOMAIN	5694	5732	EGF-LIKE 149.
FT	DOMAIN	5732	5770	EGF-LIKE 150.
FT	DOMAIN	5770	5808	EGF-LIKE 151.
FT	DOMAIN	5808	5846	EGF-LIKE 152.
FT	DOMAIN	5846	5884	EGF-LIKE 153.
FT	DOMAIN	5884	5922	EGF-LIKE 154.
FT	DOMAIN	5922	5960	EGF-LIKE 155.
FT	DOMAIN	5960	5998	EGF-LIKE 156.
FT	DOMAIN	5998	6036	EGF-LIKE 157.
FT	DOMAIN	6036	6074	EGF-LIKE 158.
FT	DOMAIN	6074	6112	EGF-LIKE 159.
FT	DOMAIN	6112	6150	EGF-LIKE 160.
FT	DOMAIN	6150	6188	EGF-LIKE 161.
FT	DOMAIN	6188	6226	EGF-LIKE 162.
FT	DOMAIN	6226	6264	EGF-LIKE 163.
FT	DOMAIN	6264	6302	EGF-LIKE 164.
FT	DOMAIN	6302	6340	EGF-LIKE 165.
FT	DOMAIN	6340	6378	EGF-LIKE 166.
FT	DOMAIN	6378	6416	EGF-LIKE 167.
FT	DOMAIN	6416	6454	EGF-LIKE 168.
FT	DOMAIN	6454	6492	EGF-LIKE 169.
FT	DOMAIN	6492	6530	EGF-LIKE 170.
FT	DOMAIN	6530	6568	EGF-LIKE 171.
FT	DOMAIN	6568	6606	EGF-LIKE 172.
FT	DOMAIN	6606	6644	EGF-LIKE 173.
FT	DOMAIN	6644	6682	EGF-LIKE 174.
FT	DOMAIN	6682	6720	EGF-LIKE 175.
FT	DOMAIN	6720	6758	EGF-LIKE 176.
FT	DOMAIN	6758	6796	EGF-LIKE 177.
FT	DOMAIN	6796	6834	EGF-LIKE 178.
FT	DOMAIN	6834	6872	EGF-LIKE 179.
FT	DOMAIN	6872	6910	EGF-LIKE 180.
FT	DOMAIN	6910	6948	EGF-LIKE 181.
FT	DOMAIN	6948	6986	EGF-LIKE 182.
FT	DOMAIN	6986	7024	EGF-LIKE 183.
FT	DOMAIN	7024	7062	EGF-LIKE 184.
FT	DOMAIN	7062	7100	EGF-LIKE 185.
FT	DOMAIN	7100	7138	EGF-LIKE 186.
FT	DOMAIN	7138	7176	EGF-LIKE 187.
FT	DOMAIN	7176	7214	EGF-LIKE 188.
FT	DOMAIN	7214	7252	EGF-LIKE 189.
FT	DOMAIN	7252	7290	EGF-LIKE 190.
FT	DOMAIN	7290	7328	EGF-LIKE 191.
FT	DOMAIN	7328	7366	EGF-LIKE 192.
FT	DOMAIN	7366	7404	EGF-LIKE 193.
FT	DOMAIN	7404	7442	EGF-LIKE 194.
FT	DOMAIN	7442	7480	EGF-LIKE 195.
FT	DOMAIN	7480	7518	EGF-LIKE 196.
FT	DOMAIN	7518	7556	EGF-LIKE 197.
FT	DOMAIN	7556	7594	EGF-LIKE 198.
FT	DOMAIN	7594	7632	EGF-LIKE 199.
FT	DOMAIN	7632	7670	EGF-LIKE 200.
FT	DOMAIN	7670	7708	EGF-LIKE 201.
FT	DOMAIN	7708	7746	EGF-LIKE 202.
FT	DOMAIN	7746	7784	EGF-LIKE 203.
FT	DOMAIN	7784	7822	EGF-LIKE 204.
FT	DOMAIN	7822	7860	EGF-LIKE 205.
FT	DOMAIN	7860	7898	EGF-LIKE 206.
FT	DOMAIN	7898	7936	EGF-LIKE 207.
FT	DOMAIN	7936	7974	EGF-LIKE 208.
FT	DOMAIN	7974	8012	EGF-LIKE 209.
FT	DOMAIN	8012	8050	EGF-LIKE 210.
FT	DOMAIN	8050	8088	EGF-LIKE 211.
FT	DOMAIN	8088	8126	EGF-LIKE 212.
FT	DOMAIN	8126	8164	EGF-LIKE 213.
FT	DOMAIN	8164	8202	EGF-LIKE 214.
FT	DOMAIN	8202	8240	EGF-LIKE 215.
FT	DOMAIN	8240	8278	EGF-LIKE 216.
FT	DOMAIN	8278	8316	EGF-LIKE 217.
FT	DOMAIN	8316	8354	EGF-LIKE 218.
FT	DOMAIN	8354	8392	EGF-LIKE 219.
FT	DOMAIN	8392	8430	EGF-LIKE 220.
FT	DOMAIN	8430	8468	EGF-LIKE 221.
FT	DOMAIN	8468	8506	EGF-LIKE 222.
FT	DOMAIN	8506	8544	EGF-LIKE 223.
FT	DOMAIN	8544	8582	EGF-LIKE 224.
FT	DOMAIN	8582	8620	EGF-LIKE 225.
FT	DOMAIN	8620	8658	EGF-LIKE 226.
FT	DOMAIN	8658	8696	EGF-LIKE 227.
FT	DOMAIN	8696	8734	EGF-LIKE 228.
FT	DOMAIN	8734	8772	EGF-LIKE 229.
FT	DOMAIN	8772	8810	EGF-LIKE 230.
FT	DOMAIN	8810	8848	EGF-LIKE 231.
FT	DOMAIN	8848	8886	EGF-LIKE 232.
FT	DOMAIN	8886	8924	EGF-LIKE 233.
FT	DOMAIN	8924	8962	EGF-LIKE 234.
FT	DOMAIN	8962	9000	EGF-LIKE 235.
FT	DOMAIN	9000	9038	EGF-LIKE 236.
FT	DOMAIN	9038	9076	EGF-LIKE 237.
FT	DOMAIN	9076	9114	EGF-LIKE 238.
FT	DOMAIN	9114	9152	EGF-LIKE 239.
FT	DOMAIN	9152	9190	EGF-LIKE 240.
FT	DOMAIN	9190	9228	EGF-LIKE 241.
FT	DOMAIN	9228	9266	EGF-LIKE 242.
FT	DOMAIN	9266	9304	EGF-LIKE 243.
FT	DOMAIN	9304	9342	EGF-LIKE 244.
FT	DOMAIN	9342	9380	EGF-LIKE 245.
FT	DOMAIN	9380	9418	EGF-LIKE 246.
FT	DOMAIN	9418	9456	EGF-LIKE 247.
FT	DOMAIN	9456	9494	EGF-LIKE 248.
FT	DOMAIN	9494	9532	EGF-LIKE 249.
FT	DOMAIN	9532	9570	EGF-LIKE 250.
FT	DOMAIN	9570	9608	EGF-LIKE 251.
FT	DOMAIN	9608	9646	EGF-LIKE 252.
FT	DOMAIN	9646	9684	EGF-LIKE 253.
FT	DOMAIN	9684	9722	EGF-LIKE 254.
FT	DOMAIN	9722	9760	EGF-LIKE 255.
FT	DOMAIN	9760	9798	EGF-LIKE 256.
FT	DOMAIN	9798	9836	EGF-LIKE 257.
FT	DOMAIN	9836	9874	EGF-LIKE 258.
FT	DOMAIN	9874	9912	EGF-LIKE 259.
FT	DOMAIN	9912	9950	EGF-L

```

Db      200 GTP-----PRINKAT-----INFSKLD-----VQYSDNPLPIF 227
QY      332 SLRSTIERHAPWVNFYITNGOIPSMINDNPRVITVTHQDVFRLSHLPTESS----- 385
Db      228 STMGNOKEHF-----ROIPTITATNEK-----THEIVRKNLKKSPYSGITGIG 273
QY      386 ---PAIESHHRIBGLSQKFIYNDVWFGKDVMPDDFYS-----HS-KG-QK 428
Db      274 PRYCSIEDKIVRFSDBRNAHQIFLEPEGHDIETYPNGISTSLPEDVQVEMTHSIGLER 333
QY      429 VYLTWFPVNCAGCGGSMWKDGYCDKACNNSACDWDGDCSGNSGSGRYTAG---GGGT 484
Db      334 AQIT-----RQYVAYEYDQCDPRLKLTLE-----SKPIEGFPLAQOIN 372
QY      485 SSGVGHFWQFG--GAINSVSZYNOGCANSWLADKFCDAQNLSCGFDAGDCGDHFHE 542
Db      373 GTTGEBEAAAGGLAGLNASLYASNKC--GWFPNR-----GQ----- 407
QY      543 LYKVILLPNQTHYIIPKGECLPYFSFAEVAKGVEGAYSDNPIRPHASIANKWTIHLIM 602
Db      408 AYLGVLIDB---LCTKGTKEPEYRMTARAEHRLI-LREDNADLRLTNIA--KSMNLID 459
QY      603 HSG-----NMATTIHFNLTFOONTNDEEFMOITVEVDTRBEGKLN--STAOKGY 649
Db      460 NSRWTRRYVEKLSNKNETRELENLKIRS-----KLVSITELANNFSSIKINTESTAK--- 510
QY      650 ENLVSPITLLPEAEILFEDIPKEKRF-PKFKSHDVNSTRAQOEYKIPLVNISLLPKDAQ 708
Db      511 -----DLTKRPEINYSTLMLEFKKSPGK---DKEAYEQIEIOEKYCGYIKROIK 557
QY      709 LSLNLTLDLQLEHGDITLKGYNISKSALRSFLMNSQHAKIKNOAITDETNDSLVAPQEK 768
Db      558 ATKV---OLNNDYIVL-----SKIKYKVVKGLSNEVV---SKL 590
QY      769 OVHKSILPNSLGVSERLQRLTPPAVSV 795
Db      591 NPYK--PYSLGQASRISGITPAISI 614

```

Search completed: July 26, 2004, 11:09:43
 Job time : 19.4568 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:05:03 ; Search time 60.6381 Seconds
(without alignments)
4828.666 Million cell updates/sec

Title: US-10-657-280-1

Perfect score: 4907
Sequence: 1 MFKLLQRYTCLSHRYGL.....SLKQLAYFTDSKNTGRQLK 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_plant: *
10: sp_ricet: *
11: sp_ricet: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4473	91.2	847	4	Q96N13
2	3314	67.5	950	4	Q9U1L2
3	837	17.1	490	4	Q8ETQ2
4	734.5	15.0	384	11	Q6I1340
5	488	9.9	666	5	Q8SX14
6	488	9.9	666	5	Q9V553
7	360	7.3	71	4	Q9HUA5
8	329	6.7	602	16	Q9L112
9	324.5	6.6	586	16	Q9L114
10	322.5	6.6	586	16	Q9S853
11	320.5	6.5	486	2	Q7X4S1
12	314	6.4	541	16	Q69851
13	287	5.8	238	2	Q9EVX1
14	287	5.8	238	2	Q8GPD3
15	266	5.4	545	2	Q84D00
16	266	5.4	545	2	Q83U59

17	266	5.4	545	16	Q9JWM8	Q9JWM8 neisseria m
18	265	5.4	545	2	Q68215	Q68215 neisseria m
19	260.5	5.3	442	5	Q861W6	Q861W6 dictyostell
20	259	5.3	545	2	Q84C29	Q84C29 neisseria m
21	258	5.3	333	2	Q8GF72	Q8GF72 streptococc
22	258	5.3	483	5	Q86RM0	Q86RM0 dictyostell
23	256	5.2	532	16	Q06628	Q06628 mycobacteri
24	256	5.2	532	16	Q7U184	Q7U184 mycobacteri
25	255.5	5.2	542	16	Q50025	Q50025 mycobacteri
26	255	5.2	370	2	Q84CH1	Q84CH1 actinobacil
27	254	5.2	364	2	Q8XSB4	Q8XSB4 actinobacil
28	252	5.1	397	5	Q86HR4	Q86HR4 dictyostell
29	249	5.0	337	16	Q88XJ7	Q88XJ7 lactobacill
30	243	5.0	333	2	Q83YR8	Q83YR8 streptococc
31	231.5	4.7	364	2	Q84ER7	Q84ER7 aeromonas h
32	230.5	4.7	366	2	Q9RG80	Q9RG80 neisseria m
33	228.5	4.7	373	2	Q51151	Q51151 neisseria m
34	226	4.6	360	2	Q84EK9	Q84EK9 aeromonas h
35	191	3.9	1476	13	Q90285	Q90285 carassius a
36	183	3.7	638	13	Q42372	Q42372 brachydanio
37	175.5	3.6	1455	5	Q86FJ9	Q86FJ9 caenorhabdi
38	175.5	3.6	2352	5	Q61240	Q61240 halocynthia
39	175	3.6	3869	5	Q86PQ3	Q86PQ3 cryptospori
40	173	3.5	2524	5	Q9GP85	Q9GP85 brachydanio
41	170.5	3.5	762	13	Q42373	Q42373 brachydanio
42	170.5	3.5	2468	13	Q800E4	Q800E4 brachydanio
43	168.5	3.4	523	5	Q9N833	Q9N833 leishmania
44	167	3.4	6118	5	Q81396	Q81396 plasmodium
45	161.5	3.3	2528	13	Q8AXP0	Q8AXP0 cypops pyrr

ALIGNMENTS

RESULT 1
ID Q96N13 PRELIMINARY; PRT; 847 AA.

AC Q96N13; 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ15175 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RA Niromiya K., Nagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsura N., Sato K., Tanikawa Y., Yamazaki Y., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto U., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Iwagai T.,
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK056137; BAB71102.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00066; notch; 1.
DR SMART; SM00004; NU; 2.
KW Hypothetical protein.
FT NON_TER 847
SQ SEQUENCE 847 AA; 95165 MW; 0A115015824733C5 CRC64;

Query Match 91.2%; Score 4473; DB 4; Length 847;
Best Local Similarity 99.8%; Pred. No. 7,9e-289;
Matches 845; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

CY 1 MFKLLQRYTCLSHRYGLVYVTVIVSAFOGFEVLEWMSKRDQHYVLFDSYRNDI 60

```

Db      1 MFLLROROTYCLSHRYGLYVCFGLVVTIVSARFQFGEVLEMSRDOYHVLFPDSYRDN1
QY      61 AAKSPONRLCLPMPIIDVYTTWNGTDLLELKLQOVESOMEBOKAMEIIGXKTTBPTK 120
Db      61 AAKSPONRLCLPMPIIDVYTTWNGTDLLELKLQOVESOMEBOKAMEIIGXKTTBPTK 120
QY      121 KSEKOLECLTHCICIVPMVLDPALPANITLKDVESLYPSFHASDIFNVAKPKNPSTNV 180
Db      121 KSEKOLECLTHCICIVPMVLDPALPANITLKDVESLYPSFHASDIFNVAKPKNPSTNV 180
QY      181 SYVVPDSTVDVDAHSGILKSGSRQVMTTDEKVEPGVIMODLFLSGPPTPET 240
Db      181 SYVVPDSTVDVDAHSGILKSGSRQVMTTDEKVEPGVIMODLFLSGPPTPET 240
QY      241 NOLKTLKPENLSSKVKLLQLYSEASVALLKLNPPDOELNKOTKONTIDKELTISPA 300
Db      241 NOLKTLKPENLSSKVKLLQLYSEASVALLKLNPPDOELNKOTKONTIDKELTISPA 300
QY      301 YLLMPLSAISQKQEDISASRFEDNEELRYSLSRIEHAAPVWRIPIVTNGQIPSWLNL 360
Db      301 YLLMPLSAISQKQEDISASRFEDNEELRYSLSRIEHAAPVWRIPIVTNGQIPSWLNL 360
QY      361 DNPRTYITVHOVFNLSHLPTFSSPAIESHIREGLSOKFIYLDVDMGKXVMPDDF 420
Db      361 DNPRTYITVHOVFNLSHLPTFSSPAIESHIREGLSOKFIYLDVDMGKXVMPDDF 420
QY      421 YHSHSGQKXYLTWPVPCNCAEGCGSWIKDGYCDKACNNACDMGDCGSGSGSRYIAG 480
Db      421 YHSHSGQKXYLTWPVPCNCAEGCGSWIKDGYCDKACNNACDMGDCGSGSGSRYIAG 480
QY      481 GGGTSGIGVHPWQGGGINSYSCNOCGANSWLADKFCDOACNVLSGCFPAGCGGDHF 540
Db      481 GGGTSGIGVHPWQGGGINSYSCNOCGANSWLADKFCDOACNVLSGCFPAGCGGDHF 540
QY      541 HELYKVIILPNQTHYIIPKGECLPYFSPAIVAKRGVEGAYSNDPIIRHASIANKKTITL 600
Db      541 HELYKVIILPNQTHYIIPKGECLPYFSPAIVAKRGVEGAYSNDPIIRHASIANKKTITL 600
QY      601 INHSGNATTHFNLTFOUNTDEEFKMQITVEVDREBPKLNSTOKXENLVSPITLLP 660
Db      601 INHSGNATTHFNLTFOUNTDEEFKMQITVEVDREBPKLNSTOKXENLVSPITLLP 660
QY      661 EAEILFEDIPKEKRPFKRHVDNSTRAQEEVKIPLVNISLPPDOALSINTLDLQEH 720
Db      661 EAEILFEDIPKEKRPFKRHVDNSTRAQEEVKIPLVNISLPPDOALSINTLDLQEH 720
QY      721 GDTILKGYLSKLSALIRSLFNMNSQAKIKNOAIITDENNDLSVAPQEKVHKSILPNSLG 780
Db      721 GDTILKGYLSKLSALIRSLFNMNSQAKIKNOAIITDENNDLSVAPQEKVHKSILPNSLG 780
QY      781 VSERLQRLTFPAVSVKVNCHDQGNPDLLETTAFRVEHTOKTIGANTKEKPSLIIV 840
Db      781 VSERLQRLTFPAVSVKVNCHDQGNPDLLETTAFRVEHTOKTIGANTKEKPSLIIV 840
QY      841 PLESQMT 847
Db      841 PLESQMT 847

RESULT 2
Q9ULI2 PRELIMINARY; PRT; 950 AA.
AC Q9ULI2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein KIAA1208 (Fragment).
GN KIAA1208
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=20039619; PubMed=10574462;
RA      Nagase T., Ishikawa K., Kikuno R., Hikosawa M., Nomura N., Ohara O.,
RT      "Prediction of the coding sequences of unidentified human genes. XV.
RT      The complete sequences of 100 new cDNA clones from brain which code
RT      for large proteins in vitro."
RL      DNA Res. 6:337-345 (1999).
DR      EMBL; AB033034; BA086522.2; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005509; F:calcium ion binding; IEA.
DR      GO; GO:0030154; P:cell differentiation; IEA.
DR      InterPro; IPR002048; EF-hand.
DR      InterPro; IPR000800; Notch_dom.
DR      Pfam; PF00036; ethand; 1.
DR      Pfam; PF00066; notch; 1.
DR      SMART; SM00004; NL; 2.
DR      PROSITE; PS00018; EF_HAND; 1.
KM      Hypothetical protein_1.
FT      NON_TER
SQ      SEQUENCE 950 AA; 108785 MW; F653BDFAEACT7503 CRC64;

Query Match 67.5%; Score 3314; DB 4; Length 950;
Best Local Similarity 99.8%; Pred. No. 1,2e-211;
Matches 621; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      307 SAISQKQEDISASRFEDNEELRYSLSRIEHAAPVWRIPIVTNGQIPSWLNLNPRVT 366
Db      1 SAISQKQEDISASRFEDNEELRYSLSRIEHAAPVWRIPIVTNGQIPSWLNLNPRVT 60
QY      367 IYTHQVFNLSHLPTFSSPAIESHIREGLSOKFIYLDVDMGKXVMPDDFYSHSGK 426
Db      61 IYTHQVFNLSHLPTFSSPAIESHIREGLSOKFIYLDVDMGKXVMPDDFYSHSGK 120
QY      427 QKXYLTWPVPCNCAEGCGSWIKDGYCDKACNNACDMGDCGSGSGSRYIAGGGGIGS 486
Db      121 QKXYLTWPVPCNCAEGCGSWIKDGYCDKACNNACDMGDCGSGSGSRYIAGGGGIGS 180
QY      487 IGVHPWQGGGINSYSCNOCGANSWLADKFCDOACNVLSGCFPAGCGGDHFHELYKV 546
Db      181 IGVHPWQGGGINSYSCNOCGANSWLADKFCDOACNVLSGCFPAGCGGDHFHELYKV 240
QY      547 ILLPNQTHYIIPKGECLPYFSPAIVAKRGVEGAYSNDPIIRHASIANKKTITLIMHSGM 606
Db      241 ILLPNQTHYIIPKGECLPYFSPAIVAKRGVEGAYSNDPIIRHASIANKKTITLIMHSGM 300
QY      607 NATTHFNLTFOUNTDEEFKMQITVEVDREBPKLNSTOKXENLVSPITLLPEAEILP 666
Db      301 NATTHFNLTFOUNTDEEFKMQITVEVDREBPKLNSTOKXENLVSPITLLPEAEILP 360
QY      667 EDIPKEKRPFKRHVDNSTRAQEEVKIPLVNISLPPDOALSINTLDLQEHGDTLK 726
Db      361 EDIPKEKRPFKRHVDNSTRAQEEVKIPLVNISLPPDOALSINTLDLQEHGDTLK 420
QY      727 GYNLSKLSALIRSLFNMNSQAKIKNOAIITDENNDLSVAPQEKVHKSILPNSLGSESLQ 786
Db      421 GYNLSKLSALIRSLFNMNSQAKIKNOAIITDENNDLSVAPQEKVHKSILPNSLGSESLQ 480
QY      787 RLTFPAVSVKVNCHDQGNPDLLETTAFRVEHTOKTIGANTKEKPSLIIVLESQM 846
Db      481 RLTFPAVSVKVNCHDQGNPDLLETTAFRVEHTOKTIGANTKEKPSLIIVLESQM 540
QY      847 TKEKKTGKREKSNMEENAHNIGTVLLGRKQIHTYDSYGLPWEKKYFQDLDLDE 906
Db      541 TKEKKTGKREKSNMEENAHNIGTVLLGRKQIHTYDSYGLPWEKKYFQDLDLDE 600
QY      907 EESLKTQLAVPTDSKNTGRQLK 928
Db      601 EESLKTQLAVPTDSKNTGRQLK 622

```

RESULT 3

Q68TQ2
ID Q68TQ2 PRELIMINARY; PRT; 490 AA.
AC Q68TQ2;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to v-naf musclicaponeurotic fibrosarcoma oncogene family,
DE protein B (Avian) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042615; AAH42615.1; -.
DR GO: GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 1.
DR PROSITE; PS00018; EF_HAND; 1.
FT NON TER 1
SQ SEQUENCE 490 AA; 57412 MW; 37B19FE0D1259AD2 CRC64;

Query Match 17.1%; Score 837; DB 4; Length 490;
Best Local Similarity 100.0%; Pred. No. 2.7e-47;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 767 EKQVHKSILPNSLGVSERLQRLTPPAVSKVNGHDQGNPPDLLETTARFEVETHQKI 826
DB 1 EKQVHKSILPNSLGVSERLQRLTPPAVSKVNGHDQGNPPDLLETTARFEVETHQKI 60
QY 827 GGNVTKKPPSLIVPLESOMTEKKITGKEKENSMEENAHNIGYTEVLIGRKLQHYTD 886
DB 61 GGNVTKKPPSLIVPLESOMTEKKITGKEKENSMEENAHNIGYTEVLIGRKLQHYTD 120
QY 887 SYLGFLPWEKKYFQDLDDEBSLKTQLAYFTDSKNTGRQK 928
DB 121 SYLGFLPWEKKYFQDLDDEBSLKTQLAYFTDSKNTGRQK 162

RESULT 4
Q68340 PRELIMINARY; PRT; 384 AA.
ID Q68340;
AC Q68340;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Basic domain/leucine zipper transcription factor (Fragment).
GN MAFB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95094266; PubMed=8001130;
RA Cordes S.F., Barch G.S.;
RT "The mouse segmentation gene kr encodes a novel basic domain-leucine
RT zipper transcription factor.";
RL Cell 79:1025-1034(1994).
DR EMBL; J36434; AAA65688.1; -.
DR PIR; I49528; I49528.
DR MGI; MGI:104555; MafB.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006357; P:regulation of transcription from pol II pro. .; IEA.
DR GO: GO:0007379; P:segment specification; IEA.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 1.
DR PROSITE; PS00018; EF_HAND; 1.
FT NON TER 1
SQ SEQUENCE 384 AA; 43754 MW; 51F473C8807A7E55 CRC64;

Query Match 15.0%; Score 734.5; DB 11; Length 384;
Best Local Similarity 57.9%; Pred. No. 1.2e-40;
Matches 165; Conservative 28; Mismatches 71; Indels 21; Gaps 6;

QY 644 TAQKCHNVLSPITLPEAELIPEDIPKPKFPKFKHVDNSTRAQOEYKLPVLNLSL 703
DB 3 TTQKAYESLVSPTPLPQADVPEDEPKERPKIRHVDNANGRFOEBVKIPRVNISL 62
QY 704 PXDAOSLNTLDQLEHGGDTLTKGYNLSSKALRSFLMNSQCHAKIKNOAIITDETDSLY 763
DB 63 PKDAQVRLSNLDQLEGGDTLTKGYNLSSKALRSFLMNSQCHAKIKNOAIITDETDSLY 121
QY 764 APOEKQVHKSILPNSLGVSERLQRLTPPAVSKVNGHDQGNPPDLLETTARFEVETHQ 823
DB 122 VPQENPSHR--PHGFAGEHRSERWTAPAVTVKGRDHALNPPVLETNARL-----AQ 174
QY 824 KTIGAVNTEKPPSLIVPLESOMTEKKITGKEKENSMEENAHNIGYTEVLIGRKLQ 883
DB 175 PTLGVTSKENSPLIVPPESHLP-----KEESDRABGNA--VPYKELPGRRCR 224
QY 884 YTDSYLGFLPWEKKYFQDLDDEBSLKTQLAYFTDSKNTGRQK 928
DB 225 IIOA--FCPGKXKXFFQDLDDEBSLKTQLAYFTDSKNTGRQK 266

RESULT 5
Q68X14 PRELIMINARY; PRT; 666 AA.
ID Q68X14;
AC Q68X14;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE RE35033P.
GN CG8027.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Drensek D., Farfan D., Frie E.,
RA George R., Gonzalez M., Guatin H., Krenniller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuncio J., Pacled J., Pargass V., Park S.,
RA Patel S., Phosnnavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceiniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY089618; AAL90356.1; -.
DR FlyBase; FBgn0033392; CG8027.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00066; notch; 1.
SQ SEQUENCE 666 AA; 77745 MW; F6FDB6D1C1C39248 CRC64;

Query Match 9.9%; Score 488; DB 5; Length 666;
Best Local Similarity 54.2%; Pred. No. 7.1e-24;
Matches 84; Conservative 30; Mismatches 41; Indels 0; Gaps 0;

QY 315 DEDISARPEDENEELYSRSIRHAPVWNTFIVNGQIPSNLNDNPRVITVHQDV 374
DB 86 DKYDSDRFDCKXELYSLSLEKXKAMIRHYIVTNGQIPSWLDSYRVTIVPREVLA 145
QY 375 RNLSHLPTRSSPAIBSHIRIBELSGKFIYLVNDVWFGSDVWPPDDYSSKQKXYVLTP 434
DB 146 PDDQQLPTSSSALEFFLRIPPLSLRFLVNDVIFLGAFLYEDLYTBAEGVRYQAWM 205
QY 435 VPNCAGCGPSWIKDGYCDKACNNSACMDGDCS 469
DB 206 VPNCALDPCWTVITIGDGAQCDRHNCVTDACGFDGDCS 240

QUERY 6
 ID 09V553 PRELIMINARY; PRT: 666 AA.
 AC 09V553;
 DT 01-MAY-2000 (TReMBUrel_13, Created)
 DT 01-JUN-2003 (TReMBUrel_25, Last sequence update)
 DT 01-OCT-2003 (TReMBUrel_24, Last annotation update)
 DE CG8027-PA.
 GN CG8027.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.J., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H., Blazer V.G., Chame N., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abri J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Baay A.A., Bakendell J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Bensn P.V., Bernick B.P., Bhandari D., Bolshakov S.,
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Duzan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dushkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Jengam C.,
 RA Uralin M., Kalush F., Karpen G.H., Ke Z., Kesterson J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P.B., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny L., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson W., Skupski M.F., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstock G.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Williams S.M., Woodaght, Worley K.C., Xu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RA Miura S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hadera S., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
 RA Clomp M.E., Drysdale R.A., Emmett D., Frisbe E., de Grey A.D.N.J.,
 RA Harris N.L., Krommiller B., Marshall B., Milburn G.H., Richter J.,
 RA Russo S., Seear S.M.J., Smith E., Shu S., Shubin G.F.,
 RA Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
 RA Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

```

RN      [4]
RP      SEQUENCE FROM N.A.
RA      FlyBase;
RL      Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AE003834; AAF58967.2; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0030154; P:cell differentiation; IEA.
DR      InterPro; IPR000800; Notch_dom.
DR      Pfam; PF00066; notch; 1.
DR      SMART; SMO0004; NL; 1.
SQ      SEQUENCE 666 AA; 77731 MW; EA232EC5C754FF6 CRC64;

Query Match          9.9%; Score 488; DB 5; Length 666;
Best Local Similarity 54.2%; Pred. No. 7,1e-24;
Matches 84; Conservative 30; Mismatches 41; Indels 0; Gaps 0;

OY      1
DB      1
OY      61 AGKSFOUR 68
DB      61 AGKSFOUR 68

RESULT 8
ID      09L112      PRELIMINARY;      PRT;      602 AA.
AC      09L112;
DT      01-OCT-2000 (TREMBlrel. 15; Created)
DT      01-OCT-2000 (TREMBlrel. 15; Last sequence update)
DT      01-MAR-2003 (TREMBlrel. 23; Last annotation update)
DE      Hypothetical protein SCO2594.
GN      SCO2594 OR SC088.05C.
OR      Streptomyces coelicolor.

```


"A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

Mo. Microbiol. 21:77-96(1996).

[14]

SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;

PX MEDLINE=21996410; PubMed=12000953;

RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H., Harper D., Bateman A., Brown S., Chanda G., Chen C.W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S., Huang C.-H., Kiser T., Larke L., Murphy L., Oliver K., O'Neill S., Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Weirzorek A., Woodward J., Barrrell B.G., Parkhill J., Hopwood D.A.;

RA "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).";

RL Nature 417:141-147(2002).

DR EMBL: AL939113; CAB55373.1; "

DR GO: 0016740; P:transferase activity; IEA.

DR GO: 0009058; P:diol synthase; IEA.

DR InterPro: IPR001296; Glyco_transf.1.

DR Pfam: PF00534; Glyco_transf.1.

DR Transferrase; Complete proteome.

SO SEQUENCE 942 AA; 105063 MW; 108B737B98A97B80 CRC64;

Query Match 6.6%; Score 324.5; DB 16; Length 942;

Best Local Similarity 28.1%; Pred. No. 8.9e-13;

Matches 93; Conservative 57; Mismatches 122; Indels 59; Gaps 11;

QY 175 NPSNTSVVVFEDSTKDYEDAHSGHLKGNRSQYRWGRGLTTDKXVPGHLMQ-----DLA 228

Db 490 NPSRLAVDAAEQTR-VKRLAAGAYEG--KAYVAILKRTPTAPGVLLERLEAVEGEVA 545

QY 229 FLSGPPTPEFTKTNQK-----TKLPENLSK-----VKLLQYSE 263

Db 546 GLRFRFRVVTSTLRFGPAVGCDIEFRWQVPEEGGDDGFVAPLRPSAVGPKLPSTLP 605

QY 264 ASVALKLNNPKQFQELNQTCK-----NMTDQKELTISPAVLMD-----LSAI 309

Db 606 AKTRV---KOREFTLEPTLRKLVSDFTPVAVTYVVDSDPRWERRRRRAALGLE 661

QY 310 SOSKQEDISASRFEDNEELRYSLSIRTERHAPWNRNFIYTNQDISMLNDPRVITV 369

Db 662 AESGGDE---AAFRPNRDELRYSLRLAMFAPIRKRLVLTDDQTPWMLNTEHGLEIVS 718

QY 370 HQDYFRLSLPTPSSPAISLHRLRLEGLSOKRTLYNDVMEFGKDYWPDFFYHSKQKY 429

Db 719 HRIPTFDQDLFPFNHSISIESQAHHDIGSEQFLYNDVDVIFGRPGAGRFPLNGASRF 778

QY 430 YLTWPVNCAGCGPGSWIXDGYCDKACNNSA 460

Db 779 F--WSPPTVVGEP-TEEDEGYFAAKNNRA 806

RESULT 10

ID 069853 PRELIMINARY; PRT; 586 AA.

AC 069853;

DT 01-AUG-1998 (TEMBLrel. 07, Created)

DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)

DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)

DS Hypothetical protein SCO6023.

GN SCO6023 OR SCIC3.11.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

CC Streptomycinae; Streptomycetaceae; Streptomyces.

CX NCBI_TaxID=1902;

[11]

SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;

PX MEDLINE=21996410; PubMed=12000953;

QY 290 IDCKELTISPAVLMDLSIISQSKODEISAREPDEBDELRYSLRSLERHAPWVREIV 3449
 Db 242 VDGND-----PAKKQKQAKGEVTHAESASAKRTISDELRYSRSLHLPAPMRINIVV 2977
 QY 350 TNGOLFSLMLNDNPRVTLYTHODVFRNLSHPPTSSPAISBHIRIFGSLQKPIYLANDV 4099
 Db 298 TDQVPAAMREDLPGARIATARELIFRANEDPTFNSSHISBQLHHIIGLLEHFEYFEDDM 3577
 QY 410 MFGKIDWPPDFYSHSKGQKVLYLWVEVN-CAEGCGPWGKIGQYCDKXCNNS 459

```
Db 358 FMGRVAPHSFT-ENGARY--FPSRKRIFOGANAE--TSPVDAACKN 403
RESULT 13
Q9EVX1 PRELIMINARY; PRT; 238 AA.
ID Q9EVX1
AC Q9EVX1
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Putative hexose transferase.
GN CPSJ.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20519245; PubMed=11065358;
RA Altmann-Roig E., Mulholland F., Gasson M.J., Griffin A.M.;
RT "The complete cps gene cluster from Streptococcus thermophilus NCPI
RT 2393 involved in the biosynthesis of a new exopolysaccharide.";
RL Microbiology 146:2793-2802(2000).
DR EMBL; Y17900; CAC18360.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
KW Transferase.
SQ SEQUENCE 238 AA; 29110 MW; A964250B01B78988 CRC64;

Query Match
Best Local Similarity 5.8%; Score 287; DB 2; Length 238;
Matches 59; Conservative 27; Mismatches 44; Indels 24; Gaps 3;

Qy 272 NNPKDFEINKOTKQMTIDGKELTISPAYLMDLSAISQSKODEDISASRFEDNEELRY 331
Db 11 NDEPIREKNTKTPHNKRD-----NDED-NVHRYDYGTFFNY 47
Qy 332 SLRSIERHAPWVNNIYITNGQIPSWLNDNPRVTIVTHQDVFRNLISHLPTSSPAIESH 391
Db 48 WFRWVERHAPWVNNIYITNGQIPSWLNVNHPKLVKVRHEE-FIPKEYLPIFNASAIEMN 106
Qy 392 IHRIEGLSOKFIYLNDDVFMGKDVWPDPPYSHSK 425
Db 107 IHRIDGISENFVLFNDMTLIDVKTSDPFVNEK 140

RESULT 14
Q8GPD3 PRELIMINARY; PRT; 238 AA.
ID Q8GPD3
AC Q8GPD3
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE EPSJ.
GN EPSJ.
OS Streptococcus thermophilus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
NCBI_TaxID=1308;
RN [1]
RP SEQUENCE FROM N.A.
RA Rallu F., Ehrlich D.S., Renaute P.;
RT "Diversity of eps operons in Streptococcus thermophilus.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF454496; AAN63705.1; -.
SQ SEQUENCE 238 AA; 29110 MW; A964250B01B78988 CRC64;

Query Match
Best Local Similarity 5.8%; Score 287; DB 2; Length 238;
Matches 59; Conservative 27; Mismatches 44; Indels 24; Gaps 3;

Qy 272 NNPKDFEINKOTKQMTIDGKELTISPAYLMDLSAISQSKODEDISASRFEDNEELRY 331
```

```
Db 11 NDEPIREKNTKTPHNKRD-----NDED-NVHRYDYGTFFNY 47
Qy 332 SLRSIERHAPWVNNIYITNGQIPSWLNDNPRVTIVTHQDVFRNLISHLPTSSPAIESH 391
Db 48 WFRWVERHAPWVNNIYITNGQIPSWLNVNHPKLVKVRHEE-FIPKEYLPIFNASAIEMN 106
Qy 392 IHRIEGLSOKFIYLNDDVFMGKDVWPDPPYSHSK 425
Db 107 IHRIDGISENFVLFNDMTLIDVKTSDPFVNEK 140

RESULT 15
Q84D00 PRELIMINARY; PRT; 545 AA.
ID Q84D00
AC Q84D00;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE SAGB.
GN SAGB.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M4775;
RA Sacchi C.T., Whitney A.A., Mayer L.W., Moherashed E., Popovic T.;
RT "Neisseria meningitidis sacb gene.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY234202; AAC85300.1; -.
SQ SEQUENCE 545 AA; 64134 MW; A3648A3C85A946F CRC64;

Query Match
Best Local Similarity 5.4%; Score 266; DB 2; Length 545;
Matches 90; Conservative 53; Mismatches 129; Indels 52; Gaps 13;

Qy 145 LPANITLKDVPSLUPSPFHSASDIFNVAKPKNPSTNVVVPDSTKQVEDAHSGLLKG-NS 203
Db 97 LPENLTKK--PALCILESHKEDFLN-----KFLTISSENKLQYKFNQIQNPKS 145
Qy 204 KQTVWNGY-----LTTDKVPEGLV-----LMQDLAFSLGFPPTKENTQGLT 245
Db 146 VNEIWTDLPSIAHVDKMLSTDRTLSSTISQFWRFECKEDKDFI---LPTARKYSR 200
Qy 246 KLPENLSKVYKQLQLYSEASVALKLNPKDFOELNKQTKQMTIDGKELTISPAYLMD 305
Db 201 KLMKH--GIKNQLFKEG-----IRNYSISSLPYEDHDNDIDLVTWNSEDKNQ 251
Qy 306 LSAISQSKD--EDISASRFEDNEELRYSLSIEHAAWVNNIYITNGQIPSWLNDN 362
Db 252 -ELYYKYPDPNSDATSTRFISRDLPALRWEMNGSFIRKIFVSNCAAPAWIDLNN 310
Qy 363 PRVTIVTHQDVFRNLISHLPTSSPAIESHIHRIEGLSOKFIYLNDDVFMGKDVWPDPPY 422
Db 311 PKIQWYHERIMQ-SALPTFSHALETSLHHPGISNFIYSNDPFLTKPLKNKNFY 369
Qy 423 HSKGQKYYL-TWPEVN--CAEGCP 443
Db 370 SNGIAKRLLEAWGNVNGECTEGEP 393
```

Search completed: July 26, 2004, 11:11:35
Job time : 63.6381 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:28 ; Search time 30.2575 Seconds

(without alignments)
3062.893 Million cell updates/sec

Title: US-10-657-280-2

Perfect score: 1731

Sequence: 1 DTFADSLRYVKNILNSKFGF.....RKIFPRRIHKASPNRIRV 328

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: Geneseqp29Jan04:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1731	100.0	328	4	ABR61377 Human GLC
2	1731	100.0	328	7	ADD27813 Human GLC
3	1731	100.0	328	7	ABW01489 Human GLC
4	1731	100.0	328	7	ABW01538 Human GLC
5	1731	100.0	1196	5	AAE25294 Human nuc
6	1731	100.0	1256	5	AAE25290 Human nuc
7	1698	98.1	328	4	ABR61380 Murine GL
8	1698	98.1	328	7	ADD27818 Human GLC
9	1698	98.1	328	7	ABW01492 Mouse pro
10	1698	98.1	328	7	ABW01541 Mouse pro
11	1487	85.9	1199	7	ABW01487 N-acetyl h
12	1487	85.9	1199	7	ABW01536 N-acetyl h
13	1487	85.9	1199	7	ABW01536 N-acetyl h
14	1487	85.9	1199	7	ABW01536 N-acetyl h
15	1487	85.9	1199	7	ABW01536 N-acetyl h
16	1487	85.9	1199	7	ABW01536 N-acetyl h
17	1487	85.9	1199	7	ABW01536 N-acetyl h
18	1487	85.9	1199	7	ABW01536 N-acetyl h
19	1487	85.9	1199	7	ABW01536 N-acetyl h
20	1487	85.9	1199	7	ABW01536 N-acetyl h
21	1487	85.9	1199	7	ABW01536 N-acetyl h
22	1487	85.9	1199	7	ABW01536 N-acetyl h
23	1487	85.9	1199	7	ABW01536 N-acetyl h
24	1487	85.9	1199	7	ABW01536 N-acetyl h
25	1487	85.9	1199	7	ABW01536 N-acetyl h

26	113.5	6.6	545	2	AAW79296
27	112	6.5	46	7	ADD27834
28	106	6.1	574	5	ABP26069
29	101.5	5.9	301	5	ABR49330
30	101.5	5.8	2633	4	ABG06505
31	101	5.8	2688	4	AAW40883
32	100.5	5.8	259	2	AAW79314
33	100.5	5.8	259	2	AAW79314
34	100.5	5.8	259	3	AAW53212
35	100.5	5.8	259	3	AAW53195
36	100.5	5.8	259	4	AAW4006
37	100.5	5.8	259	4	AAW1988
38	100.5	5.8	259	5	ABG37762
39	100.5	5.8	259	5	ABG37779
40	100.5	5.8	468	4	ABR59432
41	99.5	5.7	729	4	ABG19268
42	99.5	5.7	893	7	ADW65350
43	99.5	5.7	893	7	ADW65350
44	99.5	5.7	2202	5	ABG37482
45	99	5.7	1963	4	AAW79838

ALIGNMENTS

RESULT 1
ID ABR61377 standard; protein; 328 AA.
XX ABR61377;
AC
XX
DT 01-AUG-2003 (first entry)
XX
XX Human GLcNAc-phosphotransferase beta-subunit.
DE
XX Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;
KW GLcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;
KW N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase;
KM enzyme replacement therapy; phosphorylated lysosomal hydrolase;
KM lysosomal storage disease; enzyme; beta-subunit.
XX
XX Homo sapiens.
OS
XX
XX US6537785-B1.
XX
XX 25-MAR-2003.
XX
XX 10-AUG-2000; 2000US-00636077.
XX
XX 14-SEP-1999; 99US-0153831P.
XX
XX (GENZ-) GENZYME GLYCOTOLOGY RES INST INC.
XX
XX Canfield WM;
XX
XX WPI; 2001-290356/30.
XX
XX DR N-PSDB; ACC81001.
XX
XX Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.
XX
XX Disclosure; Page 28-29; 62pp; English.
XX
XX The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (II). The protein of the invention has nephrotropic activity, and may be useful in enzyme replacement therapy. A protein of the invention (II), (II) is useful for preparing a phosphorylated lysosomal hydrolase. The phosphorylated hydrolase comprising a terminal mannose-6-phosphate, is useful for creating a

CC patient suffering from a lysosomal storage disease. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 328 AA;

Query Match 100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 2.3e-160;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPANPHMIDRIYVQELQDMPEEFDKTSFHKVRS 60
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPANPHMIDRIYVQELQDMPEEFDKTSFHKVRS 60
QY 61 EDMQAFSFFYYILMSAVOPLNISQVFEVDITDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
DB 61 EDMQAFSFFYYILMSAVOPLNISQVFEVDITDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
QY 121 GHEHMLINSKMLPADITOLNNIPPTQESYDNPPLPVTKSLVTKCKPVTDKIKHAYKDX 180
DB 121 GHEHMLINSKMLPADITOLNNIPPTQESYDNPPLPVTKSLVTKCKPVTDKIKHAYKDX 180
QY 181 NKYRFEIMGEEBEIAFGMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHKAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFGMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHKAQTVKAVL 240
QY 241 RDPFESMFPILPSQFELPREYRNRFLMHGLOEMRAYRDLKFWTCVLAATLIMPTISFF 300
DB 241 RDPFESMFPILPSQFELPREYRNRFLMHGLOEMRAYRDLKFWTCVLAATLIMPTISFF 300
QY 301 AEQLIALKKRIFFRRRIHKEASPNRIRV 328
DB 301 AEQLIALKKRIFFRRRIHKEASPNRIRV 328

RESULT 2

ADD27813
ID ADD27813 standard; protein; 328 AA.

XX
AC ADD27813;

XX
DT 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase beta subunit.

XX human; protein phosphorylation; soluble GlcNAc-phosphotransferase;

XX UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.

XX OS Homo sapiens.

XX PN US2003119088-A1.

XX PD 26-JUN-2003.

XX PF 21-DEC-2001; 2001US-00023889.

XX PR 21-DEC-2001; 2001US-00023889.

XX PA (NOVA-) NOVAAZYME PHARM INC.

XX PI Canfield W, Kudo M;

XX DR WPI; 2003-801323/75.

XX N-PSDB; ADD27811.

XX PT Phosphorylating a protein for treating a patient suffering from a
XX lysosomal storage disease e.g. Fabry's disease by contacting the protein
XX with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
XX protein.

XX PS Claim 8; SEQ ID NO 5; 55pp; English.

XX CC The invention relates to a method of phosphorylating a protein comprising
XX contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-

CC acetylglucosamine) and producing a phosphorylated protein. The method is
CC useful for treating a patient suffering from a lysosomal storage disease
CC e.g. Fabry's disease. The present sequence represents the amino acid
CC sequence of the human GlcNAc-phosphotransferase beta subunit.

XX
SQ Sequence 328 AA;

Query Match 100.0%; Score 1731; DB 7; Length 328;
Best Local Similarity 100.0%; Pred. No. 2.3e-160;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPANPHMIDRIYVQELQDMPEEFDKTSFHKVRS 60
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPANPHMIDRIYVQELQDMPEEFDKTSFHKVRS 60
QY 61 EDMQAFSFFYYILMSAVOPLNISQVFEVDITDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
DB 61 EDMQAFSFFYYILMSAVOPLNISQVFEVDITDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
QY 121 GHEHMLINSKMLPADITOLNNIPPTQESYDNPPLPVTKSLVTKCKPVTDKIKHAYKDX 180
DB 121 GHEHMLINSKMLPADITOLNNIPPTQESYDNPPLPVTKSLVTKCKPVTDKIKHAYKDX 180
QY 181 NKYRFEIMGEEBEIAFGMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHKAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFGMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHKAQTVKAVL 240
QY 241 RDPFESMFPILPSQFELPREYRNRFLMHGLOEMRAYRDLKFWTCVLAATLIMPTISFF 300
DB 241 RDPFESMFPILPSQFELPREYRNRFLMHGLOEMRAYRDLKFWTCVLAATLIMPTISFF 300
QY 301 AEQLIALKKRIFFRRRIHKEASPNRIRV 328
DB 301 AEQLIALKKRIFFRRRIHKEASPNRIRV 328

RESULT 3

ABM01489
ID ABM01489 standard; protein; 328 AA.

XX
AC ABM01489;

XX
DT 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase beta subunit precursor protein.

XX Mannose glycoprotein; gene therapy; carbohydrate deficient cell;

XX lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;

XX gastrointestinal; human; enzyme; lectin resistant cell;

XX deoxymannojirimycin; kifunensine; glycosylation inhibition.

XX OS Homo sapiens.

XX PN US2003124652-A1.

XX PD 03-JUL-2003.

XX PF 21-DEC-2001; 2001US-00023889.

XX PR 21-DEC-2001; 2001US-00023889.

XX PA (NOVA-) NOVAAZYME PHARM INC.

XX PI Canfield WM;

XX DR WPI; 2003-810984/76.

XX N-PSDB; AAD62491.

XX PT Producing a high mannose glycoprotein for treating lysosomal storage
XX disease, comprises culturing the lectin resistant mammalian cell in the
XX presence of deoxymannojirimycin and kifunensine.

XX PS Claim 10; Page 18-19; 46pp; English.

XX The invention relates to a method for producing a high mannose
 CC glycoprotein. The method comprises: introducing and expressing a
 CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
 CC the cell in the presence of a lectin to obtain a lectin resistant cell;
 CC isolating the cell; culturing the cell in the presence of
 CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
 CC glycoprotein; and collecting the glycoprotein. The invention is useful in
 CC gene therapy. The method is useful for producing a high mannose
 CC glycoprotein in a complex carbohydrate deficient cell for treating
 CC lysosomal storage disease. The present sequence is human N-
 CC acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor
 CC protein
 CC
 SQ Sequence 328 AA;

Query Match 100.0%; Score 1731; DB 7; Length 328;
 Best Local Similarity 100.0%; Pred. No. 2,3e-160;
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKYPAAHPMHIDRIVMQELQDMPEEFDKTSFHKVRS 60
 DB 1 DTFADSLRYVKNKILNSKFGFTSRKYPAAHPMHIDRIVMQELQDMPEEFDKTSFHKVRS 60
 QY 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDRERITLATRIHELPISLQDLT 120
 DB 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDRERITLATRIHELPISLQDLT 120
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPDNLPVTKSLVYNCKPVTDKIHKAYKDK 180
 DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPDNLPVTKSLVYNCKPVTDKIHKAYKDK 180
 QY 181 NKYRFEMGESEIAPFKIRTNVSHVVGQLDDIRKRPKFCVCLNDINDHNHKAQTVKAVL 240
 DB 181 NKYRFEMGESEIAPFKIRTNVSHVVGQLDDIRKRPKFCVCLNDINDHNHKAQTVKAVL 240
 QY 241 RPFYSMFPIPSQFELPREYRNFLMHLEQEWRAVRDKLKFWTHCVLATLIVFTTISFF 300
 DB 241 RPFYSMFPIPSQFELPREYRNFLMHLEQEWRAVRDKLKFWTHCVLATLIVFTTISFF 300
 QY 301 AEQIALKRXKIFPRRIHKEASPNRIRV 328
 DB 301 AEQIALKRXKIFPRRIHKEASPNRIRV 328

RESULT 4
 ABW01538
 ID ABW01538 standard; protein; 328 AA.
 AC ABW01538;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human GlcNAc-phosphotransferase beta subunit precursor protein.
 XX
 KM Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
 KM N-acetylglucosamine-1-phosphotransferase; enzyme; gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 PN US2003124653-A1.
 XX
 PD 03-UTL-2003.
 XX
 XX 21-DEC-2001; 2001US-00023890.
 XX
 PR 21-DEC-2001; 2001US-00023890.
 XX
 PA (NOVA-) NOVAZYME PHARM INC.
 XX
 PI Canfield WM;
 XX
 DR WPI; 2003-810985/76.

DR N-PSDB; AAD62650.
 XX
 PT Producing a glycoprotein with reduced complex carbohydrates by culturing
 PT the lectin resistant mammalian cell expressing the glycoprotein for
 PT treating lysosomal storage disease.
 XX
 PS Claim 10; Page 18-19; 46pp; English.
 XX
 CC The present invention provides a method of producing a glycoprotein
 CC having reduced complex carbohydrates by culturing the lectin resistant
 CC mammalian cell expressing the glycoprotein. The method is useful for
 CC producing a glycoprotein with reduced complex carbohydrates for treating
 CC lysosomal storage disease. The present invention is also useful in gene
 CC therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-
 CC phosphotransferase beta subunit precursor protein
 CC
 SQ Sequence 328 AA;

Query Match 100.0%; Score 1731; DB 7; Length 328;
 Best Local Similarity 100.0%; Pred. No. 2,3e-160;
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKYPAAHPMHIDRIVMQELQDMPEEFDKTSFHKVRS 60
 DB 1 DTFADSLRYVKNKILNSKFGFTSRKYPAAHPMHIDRIVMQELQDMPEEFDKTSFHKVRS 60
 QY 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDRERITLATRIHELPISLQDLT 120
 DB 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDRERITLATRIHELPISLQDLT 120
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPDNLPVTKSLVYNCKPVTDKIHKAYKDK 180
 DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPDNLPVTKSLVYNCKPVTDKIHKAYKDK 180
 QY 181 NKYRFEMGESEIAPFKIRTNVSHVVGQLDDIRKRPKFCVCLNDINDHNHKAQTVKAVL 240
 DB 181 NKYRFEMGESEIAPFKIRTNVSHVVGQLDDIRKRPKFCVCLNDINDHNHKAQTVKAVL 240
 QY 241 RPFYSMFPIPSQFELPREYRNFLMHLEQEWRAVRDKLKFWTHCVLATLIVFTTISFF 300
 DB 241 RPFYSMFPIPSQFELPREYRNFLMHLEQEWRAVRDKLKFWTHCVLATLIVFTTISFF 300
 QY 301 AEQIALKRXKIFPRRIHKEASPNRIRV 328
 DB 301 AEQIALKRXKIFPRRIHKEASPNRIRV 328

RESULT 5
 AAE25294
 ID AAE25294 standard; protein; 1196 AA.
 AC AAE25294;
 XX
 DT 30-OCT-2002 (first entry)
 XX
 DE Human nucleic acid-associated protein (NAAP-13).
 XX
 KM Human; nucleic acid-associated protein; NAAP-13; neurological disorder;
 KM arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;
 KM lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
 KM autoimmune disorder; AIDS; allergy; anemia; stroke; malaria; leishmania;
 KM gene therapy; neurotropic; neuroprotective; cerebroprotective; virocidic;
 KM immunosuppressive; protozoacide; antimicrobial.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FH Peptide
 FT 1. 34
 FT /label= Signal_peptide
 FT 1. 19
 FT /note= "Cytosolic domain"
 FT 20. 42
 FT /note= "Transmembrane domain"

FT Protein 35. .1196
 FT /note="Mature human NAAP-13"
 FT Domain 43. .1152
 FT /note="Non-cytosolic domain"
 FT Domain 1153. .1175
 FT /note="Transmembrane domain"
 FT Domain 1176. .1196
 FT /note="Cytosolic domain"

XX MO200250279-A2.

XX 27-JUN-2002.

XX 19-DEC-2001; 2001WO-US050256.

XX 21-DEC-2000; 2000US-0257714P.

XX 05-JAN-2001; 2001US-0260081P.

XX 16-JAN-2001; 2001US-0262302P.

XX 23-JAN-2001; 2001US-0263823P.

XX 02-FEB-2001; 2001US-0266088P.

XX 29-OCT-2001; 2001US-0348442P.

XX (INCY-) INCYTE GENOMICS INC.

XX Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;
 PI Walla NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;
 PI Borowsky ML, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;
 PI Gierzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalia AJA;
 PI Lee EA, Yang J, Gorvad AE, Emerling BM, Margulis JR, Lee SY;
 PI Swarnakar A, Reddy R;
 XX WPI; 2002-519887/55.
 DR N-PSDB; AAD41203.

XX Nucleic acid associated proteins and nucleic acids for diagnosing,
 PT treating and preventing cell proliferative (e.g. cancers), neurological
 PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).

XX Claim 68; Page 169-172; 193pp; English.

XX The invention relates to nucleic acid-associated proteins (NAAP) and
 CC nucleic acids. The nucleic acid and amino acid sequences are useful for
 CC diagnosing, treating and preventing cell proliferative e.g.
 CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological
 CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and
 CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections
 CC (e.g. malaria, or leishmania), as well as in assessing the effects of
 CC exogenous compound on the expression of nucleic acid and amino acid
 CC sequences of nucleic acid-associated proteins. The invention is useful in
 CC gene therapy. The present sequence is human NAAP-13

XX Sequence 1196 AA;

XX Query Match 100.0%; Score 1731; DB 5; Length 1196;
 XX Best Local Similarity 100.0%; Pred. No. 1.5e-159;
 XX Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFASLSRKYVNLKSKGFTSRKYPAMPHKIDIVNQEIQDNPFPEFDTSTHKKRHS 60
 DB 869 DTFASLSRKYVNLKSKGFTSRKYPAMPHKIDIVNQEIQDNPFPEFDTSTHKKRHS 928
 QY 61 EDMQFAFSFYILMSAVOPLNISQVFDEVDTPQSGVLSDRSIRTLATRIHPLSLDOLT 120
 DB 929 EDMQFAFSFYILMSAVOPLNISQVFDEVDTPQSGVLSDRSIRTLATRIHPLSLDOLT 988
 QY 121 GIEHMLINCSKMLPADITQLNNIPTQESYDYPNIPVTKSLVTNCKEPTDKIKHAYKDX 180
 DB 969 GIEHMLINCSKMLPADITQLNNIPTQESYDYPNIPVTKSLVTNCKEPTDKIKHAYKDX 1048
 QY 181 NKCFEINGEERIAKMLRINVSHVVGQLDIPKXPKFVCLNDINIDNHDAQTVAYL 240
 DB 1049 NKCFEINGEERIAKMLRINVSHVVGQLDIPKXPKFVCLNDINIDNHDAQTVAYL 1108

QY 241 RDPFESMFPIPSQPELPREYRNFILMHLEQEWRAIRDCLKPTWTCVATLMTTFISRP 300
 DB 1109 RDPFESMFPIPSQPELPREYRNFILMHLEQEWRAIRDCLKPTWTCVATLMTTFISRP 1168
 QY 301 AEQILAKRKIFPPRRRIHKEASPNRIRV 328
 DB 1169 AEQILAKRKIFPPRRRIHKEASPNRIRV 1196

XX RESULT 6

XX AAE25290
 XX ID AAE25290 standard; protein; 1256 AA.

XX AAE25290;

XX 30-OCT-2002 (first entry)

XX Human nucleic acid-associated protein (NAAP-9).

XX Human; nucleic acid-associated protein: NAAP-9; neurological disorder;
 KW arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;
 KW lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
 KW autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;
 KW gene therapy; nootropic; neuroprotective; cerebroprotective; vitruicide;
 KW immunosuppressive; protozoacide; antimicrobial.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

CC (e.g. malaria, or leishmania), as well as in assessing the effects of
CC exogenous compound on the expression of nucleic acid and amino acid
CC sequences of nucleic acid-associated proteins. The invention is useful in
CC gene therapy. The present sequence is human NAAP-9
XX

Sequence 1256 AA;

Query Match 100.0%; Score 1731; DB 5; Length 1256;
Best Local Similarity 100.0%; Pred. No. 1.6e-159;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRHS 60
DB 929 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRHS 988
QY 61 EDMQFAPSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 969 EDMQFAPSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPLSLQDLT 1048
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLNPVTKSLVTNCKEPTDKIHKAYKDK 180
DB 1049 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLNPVTKSLVTNCKEPTDKIHKAYKDK 1108
QY 181 NKYREIMGEELIAFKMIRTNVSHVVGQLDRIKMRKRVCLNDNDHNHDKAQTIVKAYL 240
DB 1109 NKYREIMGEELIAFKMIRTNVSHVVGQLDRIKMRKRVCLNDNDHNHDKAQTIVKAYL 1168
QY 241 RDPFESMFPIPSQFELPREYRNRFLLMHLEQEMRAVRDCLKFWTHCVLATLIMFTIFSFF 300
DB 1169 RDPFESMFPIPSQFELPREYRNRFLLMHLEQEMRAVRDCLKFWTHCVLATLIMFTIFSFF 1228
QY 301 AEQILALKRKIPRRRIHKEASPNRIRV 328
DB 1229 AEQILALKRKIPRRRIHKEASPNRIRV 1256

RESULT 7

ABR61380
ID ABR61380 standard; protein; 328 AA.

AC ABR61380;

DT 01-AUG-2003 (first entry)

DE Murine GlcNAc-phosphotransferase beta-subunit.

XX Mouse; N-acetylglucosamine-1-phosphotransferase; nephrotropic;
KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;
KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;
KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;
KW lysosomal storage disease; enzyme; beta-subunit.

OS Mus musculus.

PN US6537785-B1.

PD 25-MAR-2003.

PF 10-AUG-2000; 2000US-00636077.

PR 14-SEP-1999; 99US-0153831P.

PA (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.

XX Canfield WM;

DR WPI, 2001-290356/30.

XX N-PSDB; ACC81007.

PT Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.

XX Disclosure; Page 35-36; 62pp; English.

CC The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (II). The protein of the invention has
CC nephrotropic activity, and may be useful in enzyme replacement therapy. A
CC protein of the invention (I), (II) is useful for preparing a
CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase
CC comprising a terminal mannose-6-phosphate, is useful for treating a
CC patient suffering from a lysosomal storage disease. The present sequence
CC is used in the exemplification of the invention
XX

Sequence 328 AA;

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.9e-157;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRHS 60
DB 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRHS 60
QY 61 EDMQFAPSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 61 EDMQFAPSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLNPVTKSLVTNCKEPTDKIHKAYKDK 180
DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLNPVTKSLVTNCKEPTDKIHKAYKDK 180
QY 181 NKYREIMGEELIAFKMIRTNVSHVVGQLDRIKMRKRVCLNDNDHNHDKAQTIVKAYL 240
DB 181 NKYREIMGEELIAFKMIRTNVSHVVGQLDRIKMRKRVCLNDNDHNHDKAQTIVKAYL 240
QY 241 RDPFESMFPIPSQFELPREYRNRFLLMHLEQEMRAVRDCLKFWTHCVLATLIMFTIFSFF 300
DB 241 RDPFESMFPIPSQFELPREYRNRFLLMHLEQEMRAVRDCLKFWTHCVLATLIMFTIFSFF 300
QY 301 AEQILALKRKIPRRRIHKEASPNRIRV 328
DB 301 AEQILALKRKIPRRRIHKEASPNRIRV 328

RESULT 8

ADD27818
ID ADD27818 standard; protein; 328 AA.

AC ADD27818;

DT 15-JAN-2004 (first entry)

DE GlcNAc-phosphotransferase associated protein #2.

XX mouse; protein phosphorylation; soluble GlcNAc-phosphotransferase;
KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.

OS Mus musculus.

PN US2003119088-A1.

PD 26-JUN-2003.

PF 21-DEC-2001; 2001US-00023888.

PR 21-DEC-2001; 2001US-00023888.

PA (NOVA-) NOVAZYME PHARM INC.

XX Canfield W, Kudo M;

XX WPI, 2003-801323/75.

XX Phosphorylating a protein for treating a patient suffering from a
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
PT protein.
XX
XX Disclosure; SEQ ID NO 10; 55pp; English.
XX
CC The invention relates to a method of phosphorylating a protein comprising
CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
CC acetylglucosamine) and producing a phosphorylated protein. The method is
CC useful for treating a patient suffering from a lysosomal storage disease
CC e.g. Fabry's disease. The present sequence represents the amino acid
CC sequence of a GlcNAc-phosphotransferase associated protein.
XX
SQ Sequence 328 AA;
Query Match 98.1%; Score 1698; DB 7; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.9e-157;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNKLINSKFGFTSRKVPANHPMHIDRIWQELQDMPEEFDKTSFKVYHS 60
DB 1 DTFADSLRYVNKLINSKFGFTSRKVPANHPMHIDRIWQELQDMPEEFDKTSFKVYHS 60
QY 61 EDWQFAFSFYLLMSAVQPLNISQVFEVDYDQSGVLSDRERITLARIHPLSLQDLT 120
DB 61 EDWQFAFSFYLLMSAVQPLNISQVFEVDYDQSGVLSDRERITLARIHPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQSSYDNPVPTKSLVTNCKEPTDKIKHAYYDK 180
DB 121 GLEHMLINCSKMLPADITQLNNIPPTQSSYDNPVPTKSLVTNCKEPTDKIKHAYYDK 180
QY 181 NKRYFEIMGEELIAFMKIRTNVSHVVGQLDIRKPKRFVCINDNDINHNDQAVKAVL 240
DB 181 NKRYFEIMGEELIAFMKIRTNVSHVVGQLDIRKPKRFVCINDNDINHNDQAVKAVL 240
QY 241 RDFYESMPPIPSQFELPREYRNRFLMHMELQWRAVRDCLKWTHCVLATLIIFTIFSFF 300
DB 241 RDFYESMPPIPSQFELPREYRNRFLMHMELQWRAVRDCLKWTHCVLATLIIFTIFSFF 300
QY 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
DB 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
RESULT 9
ID ABO01492 standard; protein; 328 AA.
XX ABO01492;
AC ABO01492;
XX
DT 15-JAN-2004 (first entry)
XX
DE Mouse protein #2 used to illustrate the method of the invention.
XX
KW Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
XX lysosomal storage disease; gastrointestinal; mouse;
KW lectin resistant cell; deoxymannojirimycin; kifunensine;
XX glycosylation inhibition.
XX
OS Mus musculus.
XX
PN US2003124652-A1.
XX
PD 03-JUL-2003.
XX
XX 21-DEC-2001; 2001US-00023889.
XX
XX 21-DEC-2001; 2001US-00023889.
XX
PA (NOVA-) NOVAZYME PHARM INC.

PI Canfield WM;
XX
XX WPI; 2003-810984/76.
XX
XX Producing a high mannose glycoprotein for treating lysosomal storage
PT disease comprises culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
XX
XX Disclosure; Page 25-26; 46pp; English.
XX
PS The invention relates to a method for producing a high mannose
CC glycoprotein. The method comprises: introducing and expressing a
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
CC the cell in the presence of a lectin to obtain a lectin resistant cell;
CC isolating the cell; culturing the cell in the presence of
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
CC glycoprotein; and collecting the glycoprotein. The invention is useful in
CC gene therapy. The method is useful for producing a high mannose
CC glycoprotein in a complex carbohydrate deficient cell for treating
CC lysosomal storage disease. The present sequence is mouse protein used to
CC illustrate the method of the invention
XX
SQ Sequence 328 AA;
Query Match 98.1%; Score 1698; DB 7; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.9e-157;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNKLINSKFGFTSRKVPANHPMHIDRIWQELQDMPEEFDKTSFKVYHS 60
DB 1 DTFADSLRYVNKLINSKFGFTSRKVPANHPMHIDRIWQELQDMPEEFDKTSFKVYHS 60
QY 61 EDWQFAFSFYLLMSAVQPLNISQVFEVDYDQSGVLSDRERITLARIHPLSLQDLT 120
DB 61 EDWQFAFSFYLLMSAVQPLNISQVFEVDYDQSGVLSDRERITLARIHPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQSSYDNPVPTKSLVTNCKEPTDKIKHAYYDK 180
DB 121 GLEHMLINCSKMLPADITQLNNIPPTQSSYDNPVPTKSLVTNCKEPTDKIKHAYYDK 180
QY 181 NKRYFEIMGEELIAFMKIRTNVSHVVGQLDIRKPKRFVCINDNDINHNDQAVKAVL 240
DB 181 NKRYFEIMGEELIAFMKIRTNVSHVVGQLDIRKPKRFVCINDNDINHNDQAVKAVL 240
QY 241 RDFYESMPPIPSQFELPREYRNRFLMHMELQWRAVRDCLKWTHCVLATLIIFTIFSFF 300
DB 241 RDFYESMPPIPSQFELPREYRNRFLMHMELQWRAVRDCLKWTHCVLATLIIFTIFSFF 300
QY 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
DB 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
RESULT 10
ID ABO01541 standard; protein; 328 AA.
XX ABO01541;
AC ABO01541;
XX
DT 15-JAN-2004 (first entry)
XX
DE Mouse protein #2 used to illustrate the method of the invention.
XX
KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
XX N-acetylglucosamine-1-phosphotransferase; gene therapy; mouse.
XX
OS Mus musculus.
XX
PN US2003124653-A1.
XX
PD 03-JUL-2003.
XX
XX 21-DEC-2001; 2001US-00023890.
XX
XX 21-DEC-2001; 2001US-00023890.

XX 21-DEC-2001; 2001US-00023890.
PR (NOVA-) NOVAZYME PHARM INC.
PA
XX
XX
PI Canfield WM;
XX WPI; 2003-810995/76.
DR
XX
XX
PT Producing a glycoprotein with reduced complex carbohydrates by culturing
PT the lectin resistant mammalian cell expressing the glycoprotein for
PT treating lysosomal storage disease.
XX
PS Disclosure; Page 25-26; 46pp; English.
XX
XX The present invention provides a method of producing a glycoprotein
CC having reduced complex carbohydrates by culturing the lectin resistant
CC mammalian cell expressing the glycoprotein. The method is useful for
CC producing a glycoprotein with reduced complex carbohydrates for treating
CC lysosomal storage disease. The present invention is also useful in gene
CC therapy. The present sequence is mouse protein used to illustrate the
CC method of the invention
CC
SQ Sequence 328 AA;

Query Match 98.1%; Score 1698; DB 7; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.9e-157;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKLINSKFGFTSRKVPAAHNPMDRIVMQELQDMFPEEDKTSFHKVRS 60
DB 1 DTFADSLRYVNKLINSKFGFTSRKVPAAHNPMDRIVMQELQDMFPEEDKTSFHKVRS 60
QY 61 EDMQAFSYYIYMSAVQPLNISQVDEVDITDQSGVLSDBREITLATRIHELPLSLDIT 120
DB 61 EDMQAFSYYIYMSAVQPLNISQVDEVDITDQSGVLSDBREITLATRIHELPLSLDIT 120
QY 121 GLEHMLINCSKMLPADITQLNIPPTQESYYPNLPVTKSLVTNCKEVTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPADITQLNIPPTQESYYPNLPVTKSLVTNCKEVTDKIKHAYKDK 180
QY 181 NKXRFELMGEELIAFKMIRTNVSHVVGQLDIRKPKRFVCLINDNDHNHKAQTVKAVL 240
DB 181 NKXRFELMGEELIAFKMIRTNVSHVVGQLDIRKPKRFVCLINDNDHNHKAQTVKAVL 240
QY 241 RDEFSMFPPIPSQFELPREYRNRFILMHLEQEWRAVYRDKLKFMTGCVATLIMFTISFF 300
DB 241 RDEFSMFPPIPSQFELPREYRNRFILMHLEQEWRAVYRDKLKFMTGCVATLIMFTISFF 300
QY 301 AEQIYALKRKIFPRRRIRHKASPDRIYV 328
DB 301 AEQIYALKRKIFPRRRIRHKASPDRIYV 328

RESULT 11
ADD27810
ID ADD27810 standard; protein; 1199 AA.
XX
AC ADD27810;
XX
DT 15-JAN-2004 (first entry)
XX
XX Soluble human GlcNAc-6-phosphotransferase.
DE
XX human; protein phosphorylation; soluble GlcNAc-6-phosphotransferase;
KM UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease;
KM enzyme.
XX
OS Synthetic.
OS Homo sapiens.
XX
XX US2003119088-A1.

PD 26-JUN-2003.
XX
XX 21-DEC-2001; 2001US-00023888.
PF
XX
XX 21-DEC-2001; 2001US-00023888.
PR
XX
XX (NOVA-) NOVAZYME PHARM INC.
PA
XX
XX Canfield W, Kudo M;
PI
XX
XX WPI; 2003-801323/75.
DR
DR N-PSDB; ADD27809.
XX
XX
XX Phosphorylating a protein for treating a patient suffering from a
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
PT with a soluble GlcNAc-6-phosphotransferase and producing a phosphorylated
PT protein.
XX
PS Claim 3; SEQ ID NO 2; 55pp; English.
XX
XX The invention relates to a method of phosphorylating a protein comprising
CC contacting the protein with a soluble GlcNAc-6-phosphotransferase (UDP-N-
CC acetylglucosamine) and producing a phosphorylated protein. The method is
CC useful for treating a patient suffering from a lysosomal storage disease
CC e.g. Fabry's disease. The present sequence represents the amino acid
CC sequence of soluble human GlcNAc-6-phosphotransferase.
XX
SQ Sequence 1199 AA;

Query Match 85.9%; Score 1487; DB 7; Length 1199;
Best Local Similarity 100.0%; Pred. No. 1.1e-135;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKLINSKFGFTSRKVPAAHNPMDRIVMQELQDMFPEEDKTSFHKVRS 60
DB 919 DTFADSLRYVNKLINSKFGFTSRKVPAAHNPMDRIVMQELQDMFPEEDKTSFHKVRS 978
QY 61 EDMQAFSYYIYMSAVQPLNISQVDEVDITDQSGVLSDBREITLATRIHELPLSLDIT 120
DB 979 EDMQAFSYYIYMSAVQPLNISQVDEVDITDQSGVLSDBREITLATRIHELPLSLDIT 1038
QY 121 GLEHMLINCSKMLPADITQLNIPPTQESYYPNLPVTKSLVTNCKEVTDKIKHAYKDK 180
DB 1039 GLEHMLINCSKMLPADITQLNIPPTQESYYPNLPVTKSLVTNCKEVTDKIKHAYKDK 1098
QY 181 NKXRFELMGEELIAFKMIRTNVSHVVGQLDIRKPKRFVCLINDNDHNHKAQTVKAVL 240
DB 1099 NKXRFELMGEELIAFKMIRTNVSHVVGQLDIRKPKRFVCLINDNDHNHKAQTVKAVL 1158
QY 241 RDEFSMFPPIPSQFELPREYRNRFILMHLEQEWRAVYRDKL 281
DB 1159 RDEFSMFPPIPSQFELPREYRNRFILMHLEQEWRAVYRDKL 1199

RESULT 12
ABW01487
ID ABW01487 standard; protein; 1199 AA.
XX
AC ABW01487;
XX
DT 15-JAN-2004 (first entry)
XX
XX N-acetylglucosamine-1(GlcNAc)-6-phosphotransferase.
DE
XX Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
KM lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
KM gastrointestinal; enzyme; lectin resistant cell; deoxymannojirimycin;
KM kifunensine; glycosylation inhibition.
XX
OS Unidentified.
OS
XX
XX US2003124652-A1.

PD 03-JUL-2003.
 XX
 XX 21-DEC-2001; 2001US-00023899.
 XX
 XX 21-DEC-2001; 2001US-00023899.
 XX
 XX (NOVA-) NOVAZYME PHARM INC.
 XX
 XX Canfield WM;
 XX PI
 XX WPI; 2003-810984/76.
 DR N-PSDB; AAD62490.
 XX
 PT Producing a high mannose glycoprotein for treating lysosomal storage
 PT disease, comprises culturing the lectin resistant mammalian cell in the
 PT presence of deoxymannojirimycin and kifunensine.
 XX
 XX Claim 8; Page 10-13; 46pp; English.
 XX
 CC The invention relates to a method for producing a high mannose
 CC glycoprotein. The method comprises: introducing and expressing a
 CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
 CC the cell in the presence of a lectin to obtain a lectin resistant cell;
 CC isolating the cell; culturing the cell in the presence of
 CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
 CC glycoprotein; and collecting the glycoprotein. The invention is useful in
 CC gene therapy. The method is useful for producing a high mannose
 CC glycoprotein in a complex carbohydrate deficient cell for treating
 CC lysosomal storage disease. The present sequence is N-acetylglucosamine-1
 CC (GlcNAc)-phosphotransferase
 XX
 SQ Sequence 1199 AA;

Query Match 85.9%; Score 1487; DB 7; Length 1199;
 Best Local Similarity 100.0%; Pred. No. 1.1e-135;
 Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKLINSKFGFTSRKVPAMHFMIDRIWQELQDMPEEFDKTSFKVXHS 60
 DB 919 DTFADSLRYVNKLINSKFGFTSRKVPAMHFMIDRIWQELQDMPEEFDKTSFKVXHS 978
 QY 61 EDMQAFSFFYYLMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHPLSLDILT 120
 DB 979 EDMQAFSFFYYLMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHPLSLDILT 1038
 QY 121 GLEHMLINSKMLPADITQLNNIPPTQESYDNPVPYTKSLVTNCKPVTDKIRKAYXDK 180
 DB 1039 GLEHMLINSKMLPADITQLNNIPPTQESYDNPVPYTKSLVTNCKPVTDKIRKAYXDK 1098
 QY 181 NKYREIMGEEBEIAFKMIRTNVSHVGGQDDIRKPKRFVCCLNDINDHMKDAQTVKAVL 240
 DB 1099 NKYREIMGEEBEIAFKMIRTNVSHVGGQDDIRKPKRFVCCLNDINDHMKDAQTVKAVL 1158
 QY 241 RDFESMPFIPISQFELPREYRNRFLHMHLOQEMRAYRDKLK 281
 DB 1159 RDFESMPFIPISQFELPREYRNRFLHMHLOQEMRAYRDKLK 1199

RESULT 13
 ABW01536
 ID ABW01536 standard; protein; 1199 AA.
 XX
 XX ABW01536;
 AC
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE N-acetylglucosamine-1 (GlcNAc)-phosphotransferase protein.
 XX
 KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
 KW N-acetylglucosamine-1-phosphotransferase; gene therapy; enzyme.
 XX
 OS Unidentified.
 XX

PN US2003124653-A1.
 XX
 XX 03-JUL-2003.
 XX
 XX 21-DEC-2001; 2001US-00023890.
 PF
 XX 21-DEC-2001; 2001US-00023890.
 XX
 XX (NOVA-) NOVAZYME PHARM INC.
 XX
 XX Canfield WM;
 XX PI
 XX WPI; 2003-810984/76.
 DR N-PSDB; AAD62490.
 XX
 PT Producing a glycoprotein with reduced complex carbohydrates by culturing
 PT the lectin resistant mammalian cell expressing the glycoprotein for
 PT treating lysosomal storage disease.
 XX
 XX Claim 8; Page 10-13; 46pp; English.
 XX
 CC The present invention provides a method of producing a glycoprotein
 CC having reduced complex carbohydrates by culturing the lectin resistant
 CC mammalian cell expressing the glycoprotein. The method is useful for
 CC producing a glycoprotein with reduced complex carbohydrates for treating
 CC lysosomal storage disease. The present invention is also useful in gene
 CC therapy. The present sequence is N-acetylglucosamine (GlcNAc)-
 CC phosphotransferase protein
 XX
 SQ Sequence 1199 AA;

Query Match 85.9%; Score 1487; DB 7; Length 1199;
 Best Local Similarity 100.0%; Pred. No. 1.1e-135;
 Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKLINSKFGFTSRKVPAMHFMIDRIWQELQDMPEEFDKTSFKVXHS 60
 DB 919 DTFADSLRYVNKLINSKFGFTSRKVPAMHFMIDRIWQELQDMPEEFDKTSFKVXHS 978
 QY 61 EDMQAFSFFYYLMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHPLSLDILT 120
 DB 979 EDMQAFSFFYYLMSAVOPLNISQVFEVDVTDQSGVLSDRERTLATRIHPLSLDILT 1038
 QY 121 GLEHMLINSKMLPADITQLNNIPPTQESYDNPVPYTKSLVTNCKPVTDKIRKAYXDK 180
 DB 1039 GLEHMLINSKMLPADITQLNNIPPTQESYDNPVPYTKSLVTNCKPVTDKIRKAYXDK 1098
 QY 181 NKYREIMGEEBEIAFKMIRTNVSHVGGQDDIRKPKRFVCCLNDINDHMKDAQTVKAVL 240
 DB 1099 NKYREIMGEEBEIAFKMIRTNVSHVGGQDDIRKPKRFVCCLNDINDHMKDAQTVKAVL 1158
 QY 241 RDFESMPFIPISQFELPREYRNRFLHMHLOQEMRAYRDKLK 281
 DB 1159 RDFESMPFIPISQFELPREYRNRFLHMHLOQEMRAYRDKLK 1199

RESULT 14
 ABU07381
 ID ABU07381 standard; protein; 1459 AA.
 XX
 XX ABU07381;
 AC
 XX
 DT 28-JAN-2003 (first entry)
 XX
 DE Human protein NOV12.
 XX
 KW Human; NOVX; cardiomyopathy; atherosclerosis; cancer; hypertension;
 KW diabetes; inflammation; autoimmune disorder; allergy; blood disorder;
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
 KW immunoglobulin A nephropathy; cirrhosis; arthritis; Alzheimer's disease;
 KW Parkinson's disease; gout; infection; stroke; muscular dystrophy;
 KW epilepsy; wasting disorder; neurogenesis; cell differentiation;
 KW cell proliferation; haematopoiesis; wound healing; angiogenesis;

XX gene therapy; single nucleotide polymorphism: SNP.
XX Homo sapiens.
XX WO200285922-A2.
XX 31-OCT-2002.
XX 11-APR-2002; 2002MO-US011634.
XX 23-APR-2001; 2001US-0285748P.
XX 24-APR-2001; 2001US-0286068P.
XX 25-APR-2001; 2001US-0286292P.
XX 03-MAY-2001; 2001US-0288334P.
XX 16-MAY-2001; 2001US-0291241P.
XX 14-SEP-2001; 2001US-0322284P.
XX (CURA-) CURAGEN CORP.
XX (MILL-) MILLENNIDM PHARM INC.
XX Pena CE, Guo X, Shimkets RA, Padigaru M, Kekuda R, Szytek KA,
XX Mehrahan F, Topper UN, Malyankar UM, Wasserman S, Edinger S;
XX Smithson G, Gantner E, Komuves L;
XX WPI; 2003-058712/05.
XX N-PSDB; ABX10235.
XX New NOXV polypeptides and nucleic acids, useful for preventing or
XX treating NOXV-associated disorders, e.g. cancer, cardiomyopathy,
XX atherosclerosis or diabetes, and in chromosome mapping, tissue typing or
XX pharmacogenomics.
XX Claim 1; Page 94-95; 301pp; English.
XX The invention relates to an isolated polypeptide comprising any one of 17
XX human NOXV (1-9, 10a, 10b, 11-16) appearing as AB007369-AB007385, a
XX mature form of it, or a variant of them, where one or more residues of
XX the variant differs in not more than 15 % from the residues of the
XX sequence of them and their encoding polynucleotides appearing as ABX10223
XX -ABX10239. Also included are NOXV expression vectors, transformed cells,
XX antibodies, identifying an agent that binds to or modulates the
XX expression or activity of NOXV and screening for a modulator of activity
XX or of latency or predilection to a NOXV-associated disorder. The NOXV
XX polypeptides, polynucleotides and antibodies are useful in manufacturing
XX a medicament for treating or preventing a syndrome associated with NOXV-
XX associated disorder, such as cardiomyopathy, atherosclerosis, cancer,
XX hypertension, diabetes, inflammation, autoimmune disorders, allergies,
XX blood disorders, acquired immunodeficiency syndrome (AIDS), obesity,
XX asthma, immunoglobulin (Ig) A nephropathy, cirrhosis, arthritis,
XX Alzheimer's disease, Parkinson's disease, goitre, infections (e.g.
XX bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy, and
XX other wasting disorders associated with chronic diseases. The nucleic
XX acids and polypeptides may also be used as targets for the identification
XX of small molecules that modulate or inhibit e.g. neurogenesis, cell
XX differentiation, cell proliferation, haematopoiesis, wound healing and
XX angiogenesis, in gene therapy, in generation of antibodies that bind
XX immunospecifically to NOXV substances for use in therapeutic or
XX diagnostic methods. The nucleic acids are further used as hybridisation
XX probes, in chromosome mapping, tissue typing, preventive medicine, and
XX pharmacogenomics. The polypeptides are also useful as vaccines. The
XX present sequence represents a NOXV protein.
XX Sequence 1459 AA;
XX
XX Query Match 65.9%; Score 1140; DB 6; Length 1459;
XX Best Local Similarity 98.6%; Pred. No. 1.2e-101;
XX Matches 218; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 DFFPASLRVYKILNSKRGFTSRKVPAMPHMIDIVNOCODMPPEFDKTSFHKVHRS 60
XX DFFPASLRVYKILNSKRGFTSRKVPAMPHMIDIVNOCODMPPEFDKTSFHKVHRS 1268
XX 61 EDMQFAFSFYIYMSAVQPLNISQVFDEVDTQSGVLSDREIRLATRIHELPLSLQDLT 120

DB 1269 EDMQFAFSFYIYMSAVQPLNISQVFDEVDTQSGVLSDREIRLATRIHELPLSLQDLT 1328
QY 121 GLEHMLINCSKMLPADITOLANNIPPTQESYYDPNLPVTKSLVTWCKPVTDKIKHAKYKDK 180
DB 1329 GLEHMLINCSKMLPADITOLANNIPPTQESYYDPNLPVTKSLVTWCKPVTDKIKHAKYKDK 1386
QY 181 NKYRFEIMGEEIEAFMKIRTNVSHVVGQLDDIKRNPKEVC 221
DB 1389 NKYRFEIMGEEIEAFMKIRTNVSHVVGQLDDIKRNPKEVC 1429
RESULT 15
ID AAB93369 standard; protein; 132 AA.
AC AAB93369;
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:12517.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX
XX EPI074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INSR.
XX
XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX
XX Claim 8; SEQ ID NO 12517; 2537bp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX oligonucleotides, all of which are used in the exemplification of the

CC Present invention
XX
SQ Sequence 132 AA;

Query Match 40.7%; Score 704; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.4e-60;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	197	MRTNVSHVVGQLDDIRKNPRKFCVCLNDNHDKDAQTVKAVLRDPFYESMFPISQFEL	256
DB	1	MRTNVSHVVGQLDDIRKNPRKFCVCLNDNHDKDAQTVKAVLRDPFYESMFPISQFEL	60
QY	257	PREYRNRFLMHMELQEWRAVRDCLKFWTHCVLATLIMFTISFFFAEQLIALKRKIPRRR	316
DB	61	PREYRNRFLMHMELQEWRAVRDCLKFWTHCVLATLIMFTISFFFAEQLIALKRKIPRRR	120
QY	317	IHKASPNRIRV	328
DB	121	IHKASPNRIRV	132

Search completed: July 26, 2004, 11:09:10
Job time : 32.2575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:06:34 ; Search time 9.24536 Seconds
(without alignments)
1831.549 Million cell updates/sec

Title: US-10-657-280-2
Perfect score: 1731
Sequence: 1 DTFADSLRYVKNKILNSKGF.....RKIPRRRIHKEASPNRIRV 328
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCYUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiletest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1731	100.0	328	4	US-09-635-872A-2
2	1731	100.0	328	4	US-09-636-077A-2
3	1731	100.0	328	4	US-09-636-060C-2
4	1731	100.0	328	4	US-09-986-5552-2
5	1698	98.1	328	4	US-09-635-872A-8
6	1698	98.1	328	4	US-09-636-077A-8
7	1698	98.1	328	4	US-09-636-060C-8
8	1698	98.1	328	4	US-09-986-5552-8
9	402	23.2	502	4	US-09-635-872A-13
10	402	23.2	502	4	US-09-636-077A-13
11	402	23.2	502	4	US-09-636-060C-13
12	402	23.2	502	4	US-09-986-5552-13
13	145	8.4	28	4	US-09-635-872A-27
14	145	8.4	28	4	US-09-636-077A-27
15	145	8.4	28	4	US-09-636-060C-27
16	145	8.4	28	4	US-09-986-5552-27
17	113.5	5.6	545	4	US-08-936-107A-10
18	101	5.8	2662	4	US-09-595-684B-31
19	100.5	5.8	259	3	US-08-469-318-137
20	100.5	5.8	259	3	US-08-469-318-155
21	100.5	5.8	259	3	US-08-468-609A-137
22	100.5	5.8	259	3	US-08-468-609A-155
23	100.5	5.8	259	4	US-08-446-872A-137
24	100.5	5.8	259	4	US-08-446-872A-155
25	100.5	5.8	259	4	US-08-762-227A-137
26	100.5	5.8	259	4	US-08-762-227A-155
27	100.5	5.8	259	5	PCT-US95-01185-137

28	100.5	5.8	259	5	PCT-US95-01185-155	Sequence 155, App
29	96.5	5.6	588	4	US-09-519-232-2	Sequence 2, Appli
30	93.5	5.4	652	4	US-08-559-896B-2	Sequence 2, Appli
31	93.5	5.4	652	4	US-09-351-794A-2	Sequence 2, Appli
32	93.5	5.4	1199	3	US-09-208-742-2	Sequence 2, Appli
33	93.5	5.4	1199	4	US-09-332-295-4	Sequence 4, Appli
34	93.5	5.4	1199	4	US-09-709-979-4	Sequence 4, Appli
35	93.5	5.4	1199	4	US-10-147-268-4	Sequence 4, Appli
36	93.5	5.4	2184	4	US-09-417-485D-6	Sequence 6, Appli
37	92.5	5.3	696	4	US-09-107-532A-5205	Sequence 5205, Ap
38	92	5.3	315	4	US-09-584-568C-2	Sequence 2, Appli
39	91.5	5.3	322	3	US-08-469-318-149	Sequence 149, App
40	91.5	5.3	322	3	US-08-469-318-156	Sequence 156, App
41	91.5	5.3	322	3	US-08-469-318-157	Sequence 157, App
42	91.5	5.3	322	3	US-08-468-609A-149	Sequence 149, App
43	91.5	5.3	322	3	US-08-468-609A-156	Sequence 156, App
44	91.5	5.3	322	3	US-08-468-609A-157	Sequence 157, App
45	91.5	5.3	322	4	US-08-446-872A-149	Sequence 149, App

ALIGNMENTS

```
RESULT 1
US-09-635-872A-2
; Sequence 2, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-635-872A-2
```

Query Match 100.0%; Score 1731; DB 4; Length 328;
Best local similarity 100.0%; Pred. No. 9.7e-169;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	DTFADSLRYVKNKILNSKGF	TSRKVPANPHMIDRI	VQELDMPEEF	DKTSFKV	RHS	60
DB	1	DTFADSLRYVKNKILNSKGF	TSRKVPANPHMIDRI	VQELDMPEEF	DKTSFKV	RHS	60
QY	61	EDMQAFSYFYLYMSAVQPLNISQV	DEVDTQSGVLSDR	EIRTRIA	RIHEPL	SLQDLT	120
DB	61	EDMQAFSYFYLYMSAVQPLNISQV	DEVDTQSGVLSDR	EIRTRIA	RIHEPL	SLQDLT	120
QY	121	GLEHMTLINSKMLPATITQ	LNNIPFOESYV	PNLPVTKSLVTN	CKP	YTDKIHAKYKX	180
DB	121	GLEHMTLINSKMLPATITQ	LNNIPFOESYV	PNLPVTKSLVTN	CKP	YTDKIHAKYKX	180
QY	181	NKTRFELMGEETAFKMTIR	TVSHVVGQDDIRK	NPKREVC	INDNIDH	NKDAQV	240
DB	181	NKTRFELMGEETAFKMTIR	TVSHVVGQDDIRK	NPKREVC	INDNIDH	NKDAQV	240
QY	241	RDYFESNFPDPSQFELPRE	TRNFFLMHGLQEMRA	YRDCKKWT	HCYVATL	IMFTTFSF	300
DB	241	RDYFESNFPDPSQFELPRE	TRNFFLMHGLQEMRA	YRDCKKWT	HCYVATL	IMFTTFSF	300
QY	301	AEQILAKRKIPRRRIHKEA	SPNRI	RV			328
DB	301	AEQILAKRKIPRRRIHKEA	SPNRI	RV			328

RESULT 2

```
US-09-636-077A-2
; Sequence 2, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent version 3.0
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 328
; ORGANISM: Homo sapiens
US-09-636-077A-2

Query Match      100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 9.7e-169;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKLINSKGFSTRKVPAAHMHIDRIYVQELQDMFPEEFDKTSFHKVRS 60
DB 1 DTFADSLRYVNKLINSKGFSTRKVPAAHMHIDRIYVQELQDMFPEEFDKTSFHKVRS 60
QY 61 EDMQAFSYFYILMSAVOPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDIT 120
DB 61 EDMQAFSYFYILMSAVOPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDIT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPPLPYTKSLVNTCKEPTDKIHKAYDK 180
DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPPLPYTKSLVNTCKEPTDKIHKAYDK 180
QY 181 NKYRFEIMGEEBEIAFMIRITNSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRITNSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 241 RDFYESMFPPIPSQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTIMFTIISFF 300
DB 241 RDFYESMFPPIPSQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTIMFTIISFF 300
QY 301 AEQLIALKRIKIPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIPRRRIHKEASPNRIRV 328

RESULT 3
US-09-636-060C-2
; Sequence 2, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US00CNT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent version 3.1
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 328
; ORGANISM: Homo sapiens
US-09-636-060C-2

Query Match      100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 9.7e-169;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKLINSKGFSTRKVPAAHMHIDRIYVQELQDMFPEEFDKTSFHKVRS 60
DB 1 DTFADSLRYVNKLINSKGFSTRKVPAAHMHIDRIYVQELQDMFPEEFDKTSFHKVRS 60
```

```
DB 1 DTFADSLRYVNKLINSKGFSTRKVPAAHMHIDRIYVQELQDMFPEEFDKTSFHKVRS 60
QY 61 EDMQAFSYFYILMSAVOPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDIT 120
DB 61 EDMQAFSYFYILMSAVOPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDIT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPPLPYTKSLVNTCKEPTDKIHKAYDK 180
DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPPLPYTKSLVNTCKEPTDKIHKAYDK 180
QY 181 NKYRFEIMGEEBEIAFMIRITNSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRITNSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 241 RDFYESMFPPIPSQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTIMFTIISFF 300
DB 241 RDFYESMFPPIPSQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTIMFTIISFF 300
QY 301 AEQLIALKRIKIPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIPRRRIHKEASPNRIRV 328

RESULT 4
US-09-986-552-2
; Sequence 2, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DITV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent version 3.1
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 328
; ORGANISM: Homo sapiens
US-09-986-552-2

Query Match      100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 9.7e-169;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKLINSKGFSTRKVPAAHMHIDRIYVQELQDMFPEEFDKTSFHKVRS 60
DB 1 DTFADSLRYVNKLINSKGFSTRKVPAAHMHIDRIYVQELQDMFPEEFDKTSFHKVRS 60
QY 61 EDMQAFSYFYILMSAVOPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDIT 120
DB 61 EDMQAFSYFYILMSAVOPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDIT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPPLPYTKSLVNTCKEPTDKIHKAYDK 180
DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPPLPYTKSLVNTCKEPTDKIHKAYDK 180
QY 181 NKYRFEIMGEEBEIAFMIRITNSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRITNSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 241 RDFYESMFPPIPSQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTIMFTIISFF 300
DB 241 RDFYESMFPPIPSQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTIMFTIISFF 300
QY 301 AEQLIALKRIKIPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIPRRRIHKEASPNRIRV 328
```

RESULT 5
US-09-635-872A-8
; Sequence 8, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613050
; CURRENT APPLICATION NUMBER: US/09/635, 872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153, 831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-635-872A-8

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 2.3e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
DB 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
QY 61 EDMQAFASYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
DB 61 EDMQAFASYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
QY 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKXK 180
DB 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKXK 180
QY 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTYKAVL 240
DB 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTYKAVL 240
QY 241 RDPYESMFPPIPSQFELPREYRNRFLMHLEQEWRAVRDCLKFWTHCVLATLIIMFTISFF 300
DB 241 RDPYESMFPPIPSQFELPREYRNRFLMHLEQEWRAVRDCLKFWTHCVLATLIIMFTISFF 300
QY 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
DB 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328

RESULT 6
US-09-636-077A-8
; Sequence 8, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636, 077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153, 831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-077A-8

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 2.3e-165;

Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
DB 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
QY 61 EDMQAFASYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
DB 61 EDMQAFASYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
QY 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKXK 180
DB 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKXK 180
QY 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTYKAVL 240
DB 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTYKAVL 240
QY 241 RDPYESMFPPIPSQFELPREYRNRFLMHLEQEWRAVRDCLKFWTHCVLATLIIMFTISFF 300
DB 241 RDPYESMFPPIPSQFELPREYRNRFLMHLEQEWRAVRDCLKFWTHCVLATLIIMFTISFF 300
QY 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
DB 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328

RESULT 7
US-09-636-060C-8
; Sequence 8, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US0CONT
; CURRENT APPLICATION NUMBER: US/09/636, 060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153, 831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-060C-8

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 2.3e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
DB 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
QY 61 EDMQAFASYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
DB 61 EDMQAFASYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
QY 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKXK 180
DB 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKXK 180
QY 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTYKAVL 240
DB 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTYKAVL 240
QY 241 RDPYESMFPPIPSQFELPREYRNRFLMHLEQEWRAVRDCLKFWTHCVLATLIIMFTISFF 300
DB 241 RDPYESMFPPIPSQFELPREYRNRFLMHLEQEWRAVRDCLKFWTHCVLATLIIMFTISFF 300
QY 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
DB 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328

```
Db      301 AEOIATLKRRIKIPRRRIHKEASPDRIIV 328

RESULT 8
US-09-986-552-8
; Sequence 8, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US72DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-986-552-8

Query Match      98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 2.3e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY      1 DTFADSLRYVKNILNSKFGFTSRKYPAPHPMHIDRIWQELQDMPEEFEDKTSFHKVHS 60
Db      1 DTFADSLRYVKNILNSKFGFTSRKYPAPHPMHIDRIWQELQDMPEEFEDKTSFHKVHS 60
QY      61 EDMPAFSYFYILMSAVOPLNISQVPEVDITDQSGVLSDRERTLATRIHELPLSLQDLT 120
Db      61 EDMPAFSYFYILMSAVOPLNISQVPEVDITDQSGVLSDRERTLATRIHELPLSLQDLT 120
QY      121 GLEHMLINSKMLPADITQLNIPPTQESYDDPNLPVTKSLVTNCKEVTDKIKHAYDK 180
Db      121 GLEHMLINSKMLPADITQLNIPPTQESYDDPNLPVTKSLVTNCKEVTDKIKHAYDK 180
QY      181 NKYREIINGEERIEAFKMTIRTVSHVVGQLDIDIRKNPKKFCVCLINDINDHNDQAOTVAVL 240
Db      181 NKYREIINGEERIEAFKMTIRTVSHVVGQLDIDIRKNPKKFCVCLINDINDHNDQAOTVAVL 240
QY      241 RPEYSMPPIPSQFELPREYRRLFMHHELOEMKRAYDKKFWTHCTLATIMFTISFF 300
Db      241 RPEYSMPPIPSQFELPREYRRLFMHHELOEMKRAYDKKFWTHCTLATIMFTISFF 300
QY      301 AEOIATLKRRIKIPRRRIHKEASPDRIIV 328
Db      301 AEOIATLKRRIKIPRRRIHKEASPDRIIV 328

RESULT 9
US-09-635-872A-13
; Sequence 13, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 19613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-635-872A-13

Query Match      23.2%; Score 402; DB 4; Length 502;
Best Local Similarity 34.5%; Pred. No. 1.3e-32;
Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

QY      1 DTFADSLRYVKNILNSKFGFTSRKYPAPHPMHIDRIWQELQDMPEEFEDKTSFHKVHS 60
Db      259 DIYSHSLIATNMILNAYGFKARHVLAVHGFLLIDDIYVAMORRHHQOILDTAHOFRAP 318
QY      61 EDMPAFSYFYILMSAVOPLNISQVPEVDITDQSGVLSDRERTLATRIHELPLSLQDLT 120
Db      319 TDLQVAFAYISFLMSETKMSVBEIFDEPDIGSATWSREVRITFLTRIYQEPPLDMSAMR 378
QY      121 GLEHMLINSK---MLPADITQLNIPPTQESYDDPNLPVTKSLVTNCKEVTDKIKHA 176
Db      379 YFESEVQONCTRNIGMLKVDIVESHSL--VYREYEDSNLPITRDLVAVCPILAEALAN 436
QY      177 YKDNKPREIME--EELAFKMTIRTVSHVVGQLDIDIRKNPKKFCVCLINDINDH 229
Db      437 FAVRPKYNHVSFKRISHSNFMILSNLTVESLDRLRNPKKFCVCLINDINDAN 491

RESULT 10
US-09-636-077A-13
; Sequence 13, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-636-077A-13

Query Match      23.2%; Score 402; DB 4; Length 502;
Best Local Similarity 34.5%; Pred. No. 1.3e-32;
Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

QY      1 DTFADSLRYVKNILNSKFGFTSRKYPAPHPMHIDRIWQELQDMPEEFEDKTSFHKVHS 60
Db      259 DIYSHSLIATNMILNAYGFKARHVLAVHGFLLIDDIYVAMORRHHQOILDTAHOFRAP 318
QY      61 EDMPAFSYFYILMSAVOPLNISQVPEVDITDQSGVLSDRERTLATRIHELPLSLQDLT 120
Db      319 TDLQVAFAYISFLMSETKMSVBEIFDEPDIGSATWSREVRITFLTRIYQEPPLDMSAMR 378
QY      121 GLEHMLINSK---MLPADITQLNIPPTQESYDDPNLPVTKSLVTNCKEVTDKIKHA 176
Db      379 YFESEVQONCTRNIGMLKVDIVESHSL--VYREYEDSNLPITRDLVAVCPILAEALAN 436
QY      177 YKDNKPREIME--EELAFKMTIRTVSHVVGQLDIDIRKNPKKFCVCLINDINDH 229
Db      437 FAVRPKYNHVSFKRISHSNFMILSNLTVESLDRLRNPKKFCVCLINDINDAN 491

RESULT 11
US-09-636-060C-13
; Sequence 13, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US0CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
```

PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.1
SEQ ID NO 13
LENGTH: 502
TYPE: PRF
ORGANISM: Drosophila melanogaster
US-09-636-060C-13

Query Match 23.2%; Score 402; DB 4; Length 502;
Best Local Similarity 34.5%; Pred. No. 1.3e-32;

Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAMPHMIDRIWQELQDMPEEFDKTSFHKVHS 60
DB 259 DYSLSLATNMILNRAVGFKAHVLAHVGFLIDQIVBAQRRFHQIILDTAQRFRAP 318
QY 61 EDMQAFSYFYIYMSAVOPLNISQVDEVDTDQSGVLSDRERITLARIHEPLSLDPLT 120
DB 319 TDLQAFAYYSFLMSETKMSVEEIFDEPDIDGSAITMSDRERITLARIYQPLDWSMR 378
QY 121 GLEHMLINCSK---MLPADITQANNIPTQESYVDPLPVTKSLVYTNCKEYVDKIKHA 176
DB 379 YFEVVQCTRNLMGHLKVDYEHSTL--VYERYEDSNLPITTRDLVVRCPDLAEALAN 436
QY 177 YDKKKYREIIMG--EEIAFMKIRTNVSHVVGQLDIRKNPKFVCINDINHN 229
DB 437 FAVREKYNHVSPPKRTSHSNFMULTSNLVEVESLDRLRKPRKNCINDULAN 491

RESULT 12
US-09-986-552-13

Sequence 13, Application US/09986552
Patent No. 6670165

GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM

TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 2150890577DIV

CURRENT APPLICATION NUMBER: US/09/986,552
CURRENT FILING DATE: 2001-11-09

PRIOR APPLICATION NUMBER: 09/635,872
PRIOR FILING DATE: 2000-08-10

PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14

NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.1

SEQ ID NO 13
LENGTH: 502

TYPE: PRF
ORGANISM: Drosophila melanogaster

US-09-986-552-13

Query Match 23.2%; Score 402; DB 4; Length 502;
Best Local Similarity 34.5%; Pred. No. 1.3e-32;

Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAMPHMIDRIWQELQDMPEEFDKTSFHKVHS 60
DB 259 DYSLSLATNMILNRAVGFKAHVLAHVGFLIDQIVBAQRRFHQIILDTAQRFRAP 318
QY 61 EDMQAFSYFYIYMSAVOPLNISQVDEVDTDQSGVLSDRERITLARIHEPLSLDPLT 120
DB 319 TDLQAFAYYSFLMSETKMSVEEIFDEPDIDGSAITMSDRERITLARIYQPLDWSMR 378
QY 121 GLEHMLINCSK---MLPADITQANNIPTQESYVDPLPVTKSLVYTNCKEYVDKIKHA 176
DB 379 YFEVVQCTRNLMGHLKVDYEHSTL--VYERYEDSNLPITTRDLVVRCPDLAEALAN 436
QY 177 YDKKKYREIIMG--EEIAFMKIRTNVSHVVGQLDIRKNPKFVCINDINHN 229
DB 437 FAVREKYNHVSPPKRTSHSNFMULTSNLVEVESLDRLRKPRKNCINDULAN 491

RESULT 13
US-09-635-872A-27
Sequence 27, Application US/09635872A
Patent No. 6534300

GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM

TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 1956120S0

CURRENT APPLICATION NUMBER: US/09/635,872A
CURRENT FILING DATE: 2000-08-10

PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14

NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.0

SEQ ID NO 27
LENGTH: 28

TYPE: PRF
ORGANISM: Bos taurus

US-09-635-872A-27

Query Match 8.4%; Score 145; DB 4; Length 28;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAM 28
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAM 28

RESULT 14
US-09-636-077A-27

Sequence 27, Application US/09636077A
Patent No. 6537785

GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM

TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
FILE REFERENCE: 1956120S0

CURRENT APPLICATION NUMBER: US/09/636,077A
CURRENT FILING DATE: 2000-08-10

PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14

NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.0

SEQ ID NO 27
LENGTH: 28

TYPE: PRF
ORGANISM: Bos taurus

US-09-636-077A-27

Query Match 8.4%; Score 145; DB 4; Length 28;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAM 28
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAM 28

RESULT 15
US-09-636-060C-27

Sequence 27, Application US/09636060C
Patent No. 6642038

GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM M

TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
FILE REFERENCE: 2101190S0CONT

CURRENT APPLICATION NUMBER: US/09/636,060C
CURRENT FILING DATE: 2000-08-10

PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14

NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.1

; SEQ ID NO 27
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-636-060C-27

Query Match 8.4%; Score 145; DB 4; Length 28;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKIINSKFGFTSRKYPAH 28
|||
Db 1 DTFADSLRYVKNKIINSKFGFTSRKYPAH 28

Search completed: July 26, 2004, 11:13:19
Job time : 10.2454 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:11:44 / Search time 25.2146 Seconds

(without alignments)
4073.614 Million cell updates/sec

Title: US-10-657-280-2

Perfect score: 1731
Sequence: 1 DTFADSLRYNKNINSKFGF.....RKIFPRRIHKEASPNRIRV 328

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US50_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1731	100.0	328	9	US-09-895-072-2
2	1731	100.0	328	9	US-09-895-072-2
3	1731	100.0	328	14	US-10-023-888-5
4	1731	100.0	328	14	US-10-023-888-5
5	1731	100.0	328	14	US-10-023-889-5
6	1731	100.0	328	14	US-10-023-890-5
7	1731	100.0	328	14	US-10-024-197-5
8	1731	100.0	328	14	US-10-023-894-5
9	1731	100.0	328	14	US-10-306-686-2
10	1698	98.1	328	9	US-09-895-072-8
11	1698	98.1	328	9	US-09-895-072-8
12	1698	98.1	328	14	US-10-023-888-10
13	1698	98.1	328	14	US-10-023-889-10
14	1698	98.1	328	14	US-10-023-890-10
15	1698	98.1	328	14	US-10-024-197-10
					Sequence 10, Appl

16	1698	98.1	328	14	US-10-306-686-8	Sequence 8, Appl
17	1487	85.9	1199	14	US-10-023-888-2	Sequence 2, Appl
18	1487	85.9	1199	14	US-10-023-889-2	Sequence 2, Appl
19	1487	85.9	1199	14	US-10-023-890-2	Sequence 2, Appl
20	1487	85.9	1199	14	US-10-024-197-2	Sequence 2, Appl
21	1487	85.9	1199	14	US-10-023-894-2	Sequence 2, Appl
22	1290	74.5	248	15	US-10-120-801-90	Sequence 90, Appl
23	1140	65.9	663	15	US-10-120-801-87	Sequence 87, Appl
24	1140	65.9	1459	15	US-10-120-801-26	Sequence 26, Appl
25	581	33.6	384	15	US-10-120-801-89	Sequence 89, Appl
26	506	29.2	652	15	US-10-120-801-91	Sequence 91, Appl
27	402	23.2	502	9	US-09-895-072-13	Sequence 13, Appl
28	402	23.2	502	9	US-09-895-072-13	Sequence 13, Appl
29	402	23.2	502	14	US-10-023-888-16	Sequence 16, Appl
30	402	23.2	502	14	US-10-023-889-16	Sequence 16, Appl
31	402	23.2	502	14	US-10-023-890-16	Sequence 16, Appl
32	402	23.2	502	14	US-10-024-197-16	Sequence 16, Appl
33	402	23.2	502	14	US-10-023-894-16	Sequence 16, Appl
34	402	23.2	502	14	US-10-306-686-13	Sequence 13, Appl
35	300	17.3	55	14	US-10-023-386-30775	Sequence 30775, A
36	228	13.2	112	13	US-10-079-623-349	Sequence 349, App
37	145	8.4	28	9	US-09-895-072-27	Sequence 27, Appl
38	145	8.4	28	9	US-09-895-072-27	Sequence 27, Appl
39	145	8.4	28	14	US-10-306-686-27	Sequence 27, Appl
40	124	7.2	1312	15	US-10-369-493-1950	Sequence 1950, Ap
41	112	6.5	46	16	US-10-437-963-160369	Sequence 160369,
42	110.5	6.4	1015	16	US-10-437-963-160369	Sequence 160369,
43	104.5	6.0	937	16	US-10-437-963-160369	Sequence 160369,
44	104.5	6.0	1999	15	US-10-028-248A-107	Sequence 107, App
45	104.5	6.0	1999	15	US-10-107-782-107	Sequence 107, App

ALIGNMENTS

RESULT 1
US-09-895-072-2
Sequence 2, Application US/09895072
Patent No. US2002025550A1
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM M
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 210119USOCONT
CURRENT APPLICATION NUMBER: US/09/895,072
CURRENT FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
PRIOR APPLICATION NUMBER: US 09/635,872
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 328
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-072-2
Query Match 100.0%; Score 1731; DB 9; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DTFADSLRYNKNINSKFGFTSKRVPAHMPMDIRIWMQELQMPPEEDPKTSFKYKRS 60
Db 1 DTFADSLRYNKNINSKFGFTSKRVPAHMPMDIRIWMQELQMPPEEDPKTSFKYKRS 60
QY 61 EDQPAFSSYYVYMSAVOPNISOVFEVDYTDGSGVLSDEIRITLARTIHELPLSLQDLT 120
Db 61 EDQPAFSSYYVYVYMSAVOPNISOVFEVDYTDGSGVLSDEIRITLARTIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQIANNIPFOESYYDENLPPVYKSLVTNCKPVYDKIHAKYKOK 180
Db 121 GLEHMLINCSKMLPADITQIANNIPFOESYYDENLPPVYKSLVTNCKPVYDKIHAKYKOK 180

QY 181 NKRFEEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
DB 181 NKRFEEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
QY 241 RPFYEMFPIPSQFELPREYRNFPLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
DB 241 RPFYEMFPIPSQFELPREYRNFPLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
QY 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328
DB 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328

RESULT 2
US-09-986-552-2
; Sequence 2, Application US/09986552
; Patent No. US20020150981A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 2150890577DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-552-2

Query Match 100.0%; Score 1731; DB 9; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMFPEEFDKTSFKRVKRS 60
DB 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMFPEEFDKTSFKRVKRS 60
QY 61 EDMQFAFSFYIYLMASVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
DB 61 EDMQFAFSFYIYLMASVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPTQESYVDPNLPVTKSLVTNCKEPTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPADITQLNNIPTQESYVDPNLPVTKSLVTNCKEPTDKIKHAYKDK 180
QY 181 NKRFEEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
DB 181 NKRFEEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
QY 241 RPFYEMFPIPSQFELPREYRNFPLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
DB 241 RPFYEMFPIPSQFELPREYRNFPLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
QY 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328
DB 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328

RESULT 3
US-10-023-888-5
; Sequence 5, Application US/10023888
; Publication No. US20030119088A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: SOLUBLE GLC/NAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 203515US77
; CURRENT APPLICATION NUMBER: US/10/023,888

; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-888-5

Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMFPEEFDKTSFKRVKRS 60
DB 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMFPEEFDKTSFKRVKRS 60
QY 61 EDMQFAFSFYIYLMASVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
DB 61 EDMQFAFSFYIYLMASVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPTQESYVDPNLPVTKSLVTNCKEPTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPADITQLNNIPTQESYVDPNLPVTKSLVTNCKEPTDKIKHAYKDK 180
QY 181 NKRFEEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
DB 181 NKRFEEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
QY 241 RPFYEMFPIPSQFELPREYRNFPLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
DB 241 RPFYEMFPIPSQFELPREYRNFPLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
QY 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328
DB 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328

RESULT 4
US-10-023-889-5
; Sequence 5, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBO
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-889-5

Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMFPEEFDKTSFKRVKRS 60
DB 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMFPEEFDKTSFKRVKRS 60
QY 61 EDMQFAFSFYIYLMASVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
DB 61 EDMQFAFSFYIYLMASVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNNIPTQESYVDPNLPVTKSLVTNCKEPTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPADITQLNNIPTQESYVDPNLPVTKSLVTNCKEPTDKIKHAYKDK 180

```
QY 181 NKYRFELMGEELIAFKMIRTNVSHVVGQLDDIRKNPKKFCVCLNDNDHNHKAQTVKAVL 240
; SEQUENCE: 181
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 241 RDEYSEMFPIPSQFELPREYRNRFLLMHELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
; SEQUENCE: 241
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328
; SEQUENCE: 301
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328
; SEQUENCE: 301
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens

RESULT 5
US-10-023-890-5
; Sequence 5, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYD
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-890-5
```

```
Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DTFADSLRYNKLINSKFGFTSRKVPAMPHMIDRIWMOELQDMPEEFDKTSFHKVRS 60
; SEQUENCE: 1
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 1 DTFADSLRYNKLINSKFGFTSRKVPAMPHMIDRIWMOELQDMPEEFDKTSFHKVRS 60
; SEQUENCE: 1
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 61 EDMQAFSYFYIYMSAVOPNTISQVDEVDTQSGVLSDBREIRTLATRHEPLSLQDLT 120
; SEQUENCE: 61
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 61 EDMQAFSYFYIYMSAVOPNTISQVDEVDTQSGVLSDBREIRTLATRHEPLSLQDLT 120
; SEQUENCE: 61
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 121 GLEHMLINCSKMLPADITQANNIPTQESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDX 180
; SEQUENCE: 121
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 121 GLEHMLINCSKMLPADITQANNIPTQESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDX 180
; SEQUENCE: 121
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 181 NKYRFELMGEELIAFKMIRTNVSHVVGQLDDIRKNPKKFCVCLNDNDHNHKAQTVKAVL 240
; SEQUENCE: 181
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 181 NKYRFELMGEELIAFKMIRTNVSHVVGQLDDIRKNPKKFCVCLNDNDHNHKAQTVKAVL 240
; SEQUENCE: 181
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 241 RDEYSEMFPIPSQFELPREYRNRFLLMHELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
; SEQUENCE: 241
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 241 RDEYSEMFPIPSQFELPREYRNRFLLMHELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
; SEQUENCE: 241
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328
; SEQUENCE: 301
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328
; SEQUENCE: 301
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens

RESULT 6
US-10-024-197-5
; Sequence 5, Application US/10024197
; Publication No. US20030133924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBOSIDASE AND METHODS C
; FILE REFERENCE: 209794US0
; CURRENT APPLICATION NUMBER: US/10/024,197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-024-197-5
```

```
Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DTFADSLRYNKLINSKFGFTSRKVPAMPHMIDRIWMOELQDMPEEFDKTSFHKVRS 60
; SEQUENCE: 1
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 1 DTFADSLRYNKLINSKFGFTSRKVPAMPHMIDRIWMOELQDMPEEFDKTSFHKVRS 60
; SEQUENCE: 1
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 61 EDMQAFSYFYIYMSAVOPNTISQVDEVDTQSGVLSDBREIRTLATRHEPLSLQDLT 120
; SEQUENCE: 61
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 61 EDMQAFSYFYIYMSAVOPNTISQVDEVDTQSGVLSDBREIRTLATRHEPLSLQDLT 120
; SEQUENCE: 61
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 121 GLEHMLINCSKMLPADITQANNIPTQESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDX 180
; SEQUENCE: 121
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 121 GLEHMLINCSKMLPADITQANNIPTQESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDX 180
; SEQUENCE: 121
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 181 NKYRFELMGEELIAFKMIRTNVSHVVGQLDDIRKNPKKFCVCLNDNDHNHKAQTVKAVL 240
; SEQUENCE: 181
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 181 NKYRFELMGEELIAFKMIRTNVSHVVGQLDDIRKNPKKFCVCLNDNDHNHKAQTVKAVL 240
; SEQUENCE: 181
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 241 RDEYSEMFPIPSQFELPREYRNRFLLMHELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
; SEQUENCE: 241
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 241 RDEYSEMFPIPSQFELPREYRNRFLLMHELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
; SEQUENCE: 241
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
QY 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328
; SEQUENCE: 301
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
Db 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328
; SEQUENCE: 301
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
RESULT 7
US-10-023-894-5
; Sequence 5, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CORNFELD, Stuart
; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; FILE REFERENCE: 217139US77
; CURRENT APPLICATION NUMBER: US/10/023,894
; CURRENT FILING DATE: 2001-12-21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-894-5

Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Db 181 NKYRFIMGEETIAFMIRTNVSHVGGDDIRKPKRFVCINDINDHNDKDAQVKA VL 240
QY 241 RPFYEMFPIPSQFELPREYRNFLLMHELOEMRAYRDKLKFMTWCVALTIMFTISFF 300
Db 241 RPFYEMFPIPSQFELPREYRNFLLMHELOEMRAYRDKLKFMTWCVALTIMFTISFF 300
QY 301 AEOLIALKKKIFPRRRIRHKEASPNRIRV 328
Db 301 AEOLIALKKKIFPRRRIRHKEASPNRIRV 328

RESULT 8
US-10-306-686-2
; Sequence 2, Application US/10306686
; Publication No. US20030148460A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 230397US77DIV
; CURRENT APPLICATION NUMBER: US/10/306,686
; CURRENT FILING DATE: 2002-11-29
; PRIOR APPLICATION NUMBER: 09/636,596
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-08-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-306-686-2

Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMPHMDIRIVMQLDMFPEEFDKTSFHKVRS 60
Db 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMPHMDIRIVMQLDMFPEEFDKTSFHKVRS 60
QY 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
Db 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDPNLPVTKSLVTNCKEVTDKIKHAYKDX 180
Db 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDPNLPVTKSLVTNCKEVTDKIKHAYKDX 180
QY 181 NKYRFIMGEETIAFMIRTNVSHVGGDDIRKPKRFVCINDINDHNDKDAQVKA VL 240
Db 181 NKYRFIMGEETIAFMIRTNVSHVGGDDIRKPKRFVCINDINDHNDKDAQVKA VL 240
QY 241 RPFYEMFPIPSQFELPREYRNFLLMHELOEMRAYRDKLKFMTWCVALTIMFTISFF 300
Db 241 RPFYEMFPIPSQFELPREYRNFLLMHELOEMRAYRDKLKFMTWCVALTIMFTISFF 300
QY 301 AEOLIALKKKIFPRRRIRHKEASPNRIRV 328
Db 301 AEOLIALKKKIFPRRRIRHKEASPNRIRV 328

RESULT 9
US-09-895-072-8
; Sequence 8, Application US/09895072
; Patent No. US2002002550A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 210119US00CONT
; CURRENT APPLICATION NUMBER: US/09/895,072
; CURRENT FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/635,872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-895-072-8

Query Match 98.1%; Score 1698; DB 9; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMPHMDIRIVMQLDMFPEEFDKTSFHKVRS 60
Db 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMPHMDIRIVMQLDMFPEEFDKTSFHKVRS 60
QY 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
Db 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDPNLPVTKSLVTNCKEVTDKIKHAYKDX 180
Db 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDPNLPVTKSLVTNCKEVTDKIKHAYKDX 180
QY 181 NKYRFIMGEETIAFMIRTNVSHVGGDDIRKPKRFVCINDINDHNDKDAQVKA VL 240
Db 181 NKYRFIMGEETIAFMIRTNVSHVGGDDIRKPKRFVCINDINDHNDKDAQVKA VL 240
QY 241 RPFYEMFPIPSQFELPREYRNFLLMHELOEMRAYRDKLKFMTWCVALTIMFTISFF 300
Db 241 RPFYEMFPIPSQFELPREYRNFLLMHELOEMRAYRDKLKFMTWCVALTIMFTISFF 300
QY 301 AEOLIALKKKIFPRRRIRHKEASPNRIRV 328
Db 301 AEOLIALKKKIFPRRRIRHKEASPNRIRV 328

RESULT 10
US-09-986-552-8
; Sequence 8, Application US/09986552
; Patent No. US2002015098A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-986-552-8

Query Match 98.1%; Score 1698; DB 9; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMPHMDIRIVMQLDMFPEEFDKTSFHKVRS 60
Db 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHMPHMDIRIVMQLDMFPEEFDKTSFHKVRS 60
QY 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
Db 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120

Db 61 EDMQFAFSYFYILMSAVQPLNISQVFEHVDYDQSGVLSDBREITLATRIHDLPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNIPPTQESYDYNLPVTKSLVTNCKEPTDKIHKAYXDK 180
Db 121 GLEHMLINCSKMLPANITQLNIPPTQEAAYDYNLPVTKSLVTNCKEPTDKIHKAYXDK 180
Qy 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDLRKPKRFVCLNDINDHNDQACTVAVL 240
Db 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDLRKPKRFVCLNDINDHNDQACTVAVL 240
Qy 241 RDPFESMFPIPSQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Db 241 RDPFESMFPIPSQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Qy 301 AEQIALKRRKIFPRRRIRHKEASPDRIY 328
Db 301 AEQIALKRRKIFPRRRIRHKEASPDRIY 328

RESULT 11

US-10-023-888-10
; Sequence 10, Application US/10023888
; Publication No. US20030119088A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 20351US77
; CURRENT APPLICATION NUMBER: US/10/023,888
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-888-10

Query Match 98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRS 60
Db 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRS 60
Qy 61 EDMQFAFSYFYILMSAVQPLNISQVFEHVDYDQSGVLSDBREITLATRIHDLPLSLQDLT 120
Db 61 EDMQFAFSYFYILMSAVQPLNISQVFEHVDYDQSGVLSDBREITLATRIHDLPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNIPPTQESYDYNLPVTKSLVTNCKEPTDKIHKAYXDK 180
Db 121 GLEHMLINCSKMLPANITQLNIPPTQEAAYDYNLPVTKSLVTNCKEPTDKIHKAYXDK 180
Qy 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDLRKPKRFVCLNDINDHNDQACTVAVL 240
Db 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDLRKPKRFVCLNDINDHNDQACTVAVL 240
Qy 241 RDPFESMFPIPSQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Db 241 RDPFESMFPIPSQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Qy 301 AEQIALKRRKIFPRRRIRHKEASPDRIY 328
Db 301 AEQIALKRRKIFPRRRIRHKEASPDRIY 328

RESULT 12

US-10-023-889-10
; Sequence 10, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBOH

; TITLE OF INVENTION: DEFICIENT CELLS
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-889-10

Query Match 98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRS 60
Db 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRS 60
Qy 61 EDMQFAFSYFYILMSAVQPLNISQVFEHVDYDQSGVLSDBREITLATRIHDLPLSLQDLT 120
Db 61 EDMQFAFSYFYILMSAVQPLNISQVFEHVDYDQSGVLSDBREITLATRIHDLPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNIPPTQESYDYNLPVTKSLVTNCKEPTDKIHKAYXDK 180
Db 121 GLEHMLINCSKMLPANITQLNIPPTQEAAYDYNLPVTKSLVTNCKEPTDKIHKAYXDK 180
Qy 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDLRKPKRFVCLNDINDHNDQACTVAVL 240
Db 181 NKYRFEIMGEELIAFKMIRTNVSHVVGQLDLRKPKRFVCLNDINDHNDQACTVAVL 240
Qy 241 RDPFESMFPIPSQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Db 241 RDPFESMFPIPSQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Qy 301 AEQIALKRRKIFPRRRIRHKEASPDRIY 328
Db 301 AEQIALKRRKIFPRRRIRHKEASPDRIY 328

RESULT 13
US-10-023-890-10
; Sequence 10, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOH

; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-890-10

Query Match 98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRS 60
Db 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRS 60
Qy 61 EDMQFAFSYFYILMSAVQPLNISQVFEHVDYDQSGVLSDBREITLATRIHDLPLSLQDLT 120
Db 61 EDMQFAFSYFYILMSAVQPLNISQVFEHVDYDQSGVLSDBREITLATRIHDLPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNIPPTQESYDYNLPVTKSLVTNCKEPTDKIHKAYXDK 180

```
Db      121 GLEHMLINCSKMLPANITQLNNIPPTQEAAYDNNLPVTKSLVTNCKEVTDKIKHAYKDK 180
Qy      181 NKYRFEIMGEEBEELAFQMIRTNVSHVVGQDDIRKPKRFVCCLNDINDHNDQAOTVKA VL 240
Db      181 NKYRFEIMGEEBEELAFQMIRTNVSHVVGQDDIRKPKRFVCCLNDINDHNDQAOTVKA VL 240
Qy      241 RDFYESMFPISPQFELPREYRNRFLMHHELQEMRAYRDKLKFWTHCVLATLIMFTIFSFF 300
Db      241 RDFYESMFPISPQFELPREYRNRFLMHHELQEMRAYRDKLKFWTHCVLATLIMFTIFSFF 300
Qy      301 AEQIILAKRKIFPERRRIHKEASPDRI RV 328
Db      301 AEQIILAKRKIFPERRRIHKEASPDRI RV 328
```

```
RESULT 14
US-10-024-197-10
; Sequence 10, Application US/10024197
; Publication No. US20030133924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS
; FILE REFERENCE: 209794USO
; CURRENT APPLICATION NUMBER: US/10/024,197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-024-197-10
```

```
Query Match      98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 DTFADSLRYVKNKILNSKFGFTSRKVPANHPMHIDRIVMQELQDMPEEFDKTSFHKVRHS 60
Db      1 DTFADSLRYVKNKILNSKFGFTSRKVPANHPMHIDRIVMQELQDMPEEFDKTSFHKVRHS 60
Qy      61 EDMQFAFSFYIYMSAVOPLNISQVFEVDVTDQSGVLSDRERITLATRIHDLPLSLQDLT 120
Db      61 EDMQFAFSFYIYMSAVOPLNISQVFEVDVTDQSGVLSDRERITLATRIHDLPLSLQDLT 120
Qy      121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNNLPVTKSLVTNCKEVTDKIKHAYKDK 180
Db      121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNNLPVTKSLVTNCKEVTDKIKHAYKDK 180
Qy      181 NKYRFEIMGEEBEELAFQMIRTNVSHVVGQDDIRKPKRFVCCLNDINDHNDQAOTVKA VL 240
Db      181 NKYRFEIMGEEBEELAFQMIRTNVSHVVGQDDIRKPKRFVCCLNDINDHNDQAOTVKA VL 240
Qy      241 RDFYESMFPISPQFELPREYRNRFLMHHELQEMRAYRDKLKFWTHCVLATLIMFTIFSFF 300
Db      241 RDFYESMFPISPQFELPREYRNRFLMHHELQEMRAYRDKLKFWTHCVLATLIMFTIFSFF 300
Qy      301 AEQIILAKRKIFPERRRIHKEASPDRI RV 328
Db      301 AEQIILAKRKIFPERRRIHKEASPDRI RV 328
```

```
RESULT 15
US-10-023-894-10
; Sequence 10, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; APPLICANT: KORNFIELD, Stuart
; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; TITLE OF INVENTION: ACETYLGUCCOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSAMINIDS
```

```
; FILE REFERENCE: 217139US77
; CURRENT APPLICATION NUMBER: US/10/023,894
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-894-10
```

```
Query Match      98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 DTFADSLRYVKNKILNSKFGFTSRKVPANHPMHIDRIVMQELQDMPEEFDKTSFHKVRHS 60
Db      1 DTFADSLRYVKNKILNSKFGFTSRKVPANHPMHIDRIVMQELQDMPEEFDKTSFHKVRHS 60
Qy      61 EDMQFAFSFYIYMSAVOPLNISQVFEVDVTDQSGVLSDRERITLATRIHDLPLSLQDLT 120
Db      61 EDMQFAFSFYIYMSAVOPLNISQVFEVDVTDQSGVLSDRERITLATRIHDLPLSLQDLT 120
Qy      121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNNLPVTKSLVTNCKEVTDKIKHAYKDK 180
Db      121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNNLPVTKSLVTNCKEVTDKIKHAYKDK 180
Qy      181 NKYRFEIMGEEBEELAFQMIRTNVSHVVGQDDIRKPKRFVCCLNDINDHNDQAOTVKA VL 240
Db      181 NKYRFEIMGEEBEELAFQMIRTNVSHVVGQDDIRKPKRFVCCLNDINDHNDQAOTVKA VL 240
Qy      241 RDFYESMFPISPQFELPREYRNRFLMHHELQEMRAYRDKLKFWTHCVLATLIMFTIFSFF 300
Db      241 RDFYESMFPISPQFELPREYRNRFLMHHELQEMRAYRDKLKFWTHCVLATLIMFTIFSFF 300
Qy      301 AEQIILAKRKIFPERRRIHKEASPDRI RV 328
Db      301 AEQIILAKRKIFPERRRIHKEASPDRI RV 328
```

```
Search completed: July 26, 2004, 11:25:26
Job time : 26.2146 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:58 ; Search time 5.46316 Seconds
(without alignments)
3126.212 Million cell updates/sec

Title: US-10-657-280-2

Perfect score: 1731

Sequence: 1 DTFADSLRYVKNILNKRSGF.....RKIPRRRIKHEASPNRIKV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	124	7.2	1312	1	RA50_YEAST
2	106.5	6.2	974	1	OPAI_ONCYA
3	106	6.1	571	1	EZRA_STRPY
4	104	6.0	574	1	EZRA_STRP3
5	101	5.8	574	1	EZRA_STRP8
6	101	5.8	2663	1	CENE_HUMAN
7	99.5	5.7	587	1	TATR_NPYVM
8	99	5.7	385	1	GBA2_SOYBN
9	97	5.6	1961	1	MYH9_EAT
10	96	5.5	515	1	VAC1_YEAST
11	95	5.5	2867	1	RBP2_PLAVB
12	94	5.4	810	1	AMDY_YEAST
13	94	5.4	817	1	YGAB_YEAST
14	94	5.4	1978	1	MYH8_CHICK
15	93.5	5.4	353	1	STM_BORBU
16	93.5	5.4	1960	1	MYH9_HUMAN
17	91.5	5.3	587	1	YN4_YEAST
18	91.5	5.3	610	1	VATRA_TRYCO
19	91.5	5.3	1006	1	DPOL_VACCC
20	91.5	5.3	1006	1	DPOL_VACCC
21	91	5.3	582	1	TATR_NPVAV
22	91	5.3	1157	1	DD37_HUMAN
23	90	5.2	734	1	RELA_BACSU
24	90	5.2	920	1	CAPP_STRPY
25	89.5	5.2	460	1	YS41_CAEEL
26	89.5	5.2	1005	1	DPOL_VAVR
27	89.5	5.2	1701	1	MSP1_PLAFR
28	89.5	5.2	1701	1	MSP1_PLAFR
29	89	5.1	420	1	NCE2_RAT
30	89	5.1	879	1	RA50_SULTO
31	89	5.1	1013	1	PRML_DROME
32	89	5.1	2104	1	MYG3_SCHPO
33	88.5	5.1	463	1	POF2_SCHPO

34	88	5.1	932	1	CAPP_STRP3	Q8K873 streptococc
35	88	5.1	1118	1	CARB_YEAST	P03965 saccharomyc
36	88	5.1	2349	1	TPR_HUMAN	P12710 homo sapien
37	87.5	5.1	499	1	TPR9_CLOTE	O892X3 c biunctio
38	87.5	5.1	564	1	EZRA_STRAP	O8CWN5 staphylococ
39	87.5	5.1	661	1	HS74_TRYEB	P11145 trypanosoma
40	87.5	5.1	769	1	SMIE_KULPA	P4018 kluyveromyc
41	87.5	5.1	1526	1	MYG2_SCHPO	Q9US16 schizosacch
42	87	5.0	372	1	Y086_METJA	Q57551 methanococ
43	87	5.0	435	1	RPOD_SORBI	Q01923 sorghum bic
44	87	5.0	439	1	FGL2_HUMAN	O14314 homo sapien
45	87	5.0	576	1	PTL_BUCBP	Q8B04 buchnera ap

ALIGNMENTS

RESULT 1
RA50_YEAST STANDARD; PRT; 1312 AA.
AC P15753;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA repair protein RAD50 (153 kDa protein).
GN RAD50 OR YNL250W OR N0872.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB821;
RX MEDLINE=89276917; PubMed=2659437;
RA Alani E., Subbiah S., Kleckner N.;
RT "The yeast RAD50 gene encodes a predicted 153-kD protein containing a
RT purine nucleotide-binding domain and two large heptad-repeat
RT regions.";
RL Genetics 122:47-57(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldeaner U., Behnauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SUI1
RT from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL Yeast 13:849-860(1997).
-1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
rad50/mre11 complex possesses single-strand endonuclease activity.
and ATP-dependent double-strand-specific exonuclease activity.
RAD50 provides an ATP-dependent control of mre11 by unwinding
and/or repositioning DNA ends into the mre11 active site.
-1- SUBUNIT: Forms a complex with MRE11.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: X14814; CA932919.1; -
EMBL: X96722; CA65494.1; -
EMBL: Z71526; CA96157.1; -
PIR: S05808; BMEYDL.
GenMOnline: 143256; -
SGD: S0005194; RAD50.
GO: GO:0005634; C:nucleus, IDA.
GO: GO:0003690; F:double-stranded DNA binding, IDA.
GO: GO:0005515; F:protein binding, IPI.
GO: GO:0006003; P:double-strand break repair via nonhomologous . . ; IMP.

DR GO:000722; P:telomerase-independent telomere maintenance; IMP.
 DR InterPro: IPR003593; AAA_Arase.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR004584; Rad50.
 DR InterPro: IPR007517; Rad50_zn_hook.
 DR Pfam: PF04423; Rad50_zn_hook; 1.
 DR SMART: SM00382; AAA; 1.
 DR TIGRPFAM: TIGR00606; rad50; 1.
 DR DNA_repair; Hydrolyase; ATP-binding; Coiled coil; Meiosis.
 KM NP_BIND 34 41 ATP (POTENTIAL).
 FT DOMAIN 185 347 COILED COIL (POTENTIAL).
 FT DOMAIN 403 558 COILED COIL (POTENTIAL).
 FT DOMAIN 617 772 COILED COIL (POTENTIAL).
 FT DOMAIN 734 1108 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1312 AA; 152568 MW; 58A0A173AC5677E CRC64;
 Query Match 7.2%; Score 124; DB 1; Length 112;
 Best Local Similarity 21.3%; Pred. No. 0.17; Indels 60; Gaps 14;
 Matches 68; Conservative 65; Mismatches 126;
 QY 5 DELRYVVKILNSKFGFTSRKVPAMPHMIDR-----IWMQLQ--DMFPBFD--KTSF 54
 DB 189 DNLKSIKDMQSVDIKILKQSV-EHLKLDKDRSKAMKLNHQLQTKIDQNEEVSIEISQL 247
 QY 55 HAVRSEDMQAFASY-FYVLMASVQPLINSQVFEVDVDDQSGVSDREIRTLARIEHP 113
 DB 248 NEEKSKDLKFKSNODFKILSKVENLKNFKL-----SISD-QVKLSNSIDILD 296
 QY 114 LSLQDITGLEHMLINCSKMLPADITOLNNIPPTQESYVD--PMLPVTKSLVTNCKPYTD 171
 DB 297 LSKPD---LQNLPLANSKVLMDKNQNLRLDTISLMDRQSLGSLNSLIRRGEL-- 351
 QY 172 KIHKAYKDKKATRFELMGEELAFKQIRTNVSHV---VGQLDIRKPKRFVC--LMDN 225
 DB 352 EAGKETEKRNHLSLKE--AFQHKFQGLSINNSDMAQVNHMSQKFAFISQDLTDT 408
 QY 226 IDENHKAQTVAAVADPFESMFPISPQFELPREYRNFLE----- 266
 DB 409 IDQFAKDITLKTSLSDLSLKITVDSQNLFNKPKRSKLIHDEELAKKSPKSLSTQD 468
 QY 267 --MHLEQWRAVRDKLKFV 283
 DB 469 SLNHELENKTYEKELQSW 487
 RESULT 2
 OPAL_ONCMA STANDARD; PRT; 971 AA.
 ID OPAL_ONCMA
 AC 093248;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dymamin-like 120 kDa protein, mitochondrial precursor (120-kDa G
 protein expressed in motor neurons) (mg120).
 GN OPAL.
 OS Oncorhynchus masou (Cherry salmon) (Masu salmon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Procacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 OC NCBI_TaxID=8020;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=98372640; PubMed=9708909;
 RA Kubokawa K., Miyashita T., Kubo Y.;
 RT "Isolation of a cDNA for a novel 120-kDa GTP-binding protein expressed
 in motor neurons in the salmon brain.";
 RL FEBS Lett. 431:231-235(1998).
 CC -1- FUNCTION: May be involved in mitochondrial biogenesis.
 CC -1- SUBCELLULAR LOCATION: Mitochondrion (By similarity).
 CC -1- TISSUE SPECIFICITY: Strongly expressed in the brain, ovary and
 CC skeletal muscle. In the brain, expression of the mRNA was observed
 CC specifically in motor neurons, in nucleus oculomotorius, in
 CC nucleus valvulae lateralis, in the medulla oblongata and in the

CC spinal cord.
 CC -1- SIMILARITY: Belongs to the dymamin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
 CC or send an email to license@sb-stb.ch).
 DR EMBL: AB012720; BA032279.1; -.
 DR PIR: T00394; T00394.
 DR GO: GO:0030425; C:cytosol; ISS.
 DR GO: GO:0030061; C:mitochondrial cristae; ISS.
 DR GO: GO:0005758; C:mitochondrial intermembrane space; ISS.
 DR GO: GO:0005741; C:mitochondrial outer membrane; ISS.
 DR GO: GO:0003924; F:GTPase activity; IDA.
 DR GO: GO:0000287; F:magnesium ion binding; ISS.
 DR GO: GO:0019896; P:axon transport of mitochondrion; ISS.
 DR GO: GO:0007007; P:inner mitochondrial membrane organization a. .; ISS.
 DR GO: GO:0000266; P:mitochondrial fission; ISS.
 DR GO: GO:0008053; P:mitochondrial fusion; ISS.
 DR GO: GO:0045768; P:positive regulation of anti-apoptosis; ISS.
 DR GO: GO:0007601; P:vision; ISS.
 DR InterPro: IPR001401; Dymamin.
 DR Pfam: PF00350; dymamin; 1.
 DR PRINTS: PR00195; DYNAMIN.
 DR SMART: SM00053; DYN; 1.
 KM Motor protein; GTP-binding; Mitochondrion; Transit peptide.
 FT TRANSIT 1 971 MITOCHONDRION (POTENTIAL).
 FT NP_BIND 307 314 DYNAMIN-LIKE 120 KDA PROTEIN.
 FT NP_BIND 410 414 GTP (POTENTIAL).
 FT NP_BIND 461 461 GTP (POTENTIAL).
 SQ SEQUENCE 971 AA; 112069 MW; 462958C03C1A4F02 CRC64;
 Query Match 6.2%; Score 106.5; DB 1; Length 971;
 Best Local Similarity 23.0%; Pred. No. 2.4; Indels 85; Gaps 14;
 Matches 62; Conservative 43; Mismatches 79;
 QY 7 LRYVVKILNSKFGFTSRKVPAMPHMIDRIWMQLQDMFPBFDKTSFKHRSSEDMQ 64
 DB 102 LRYI-VLGSVAGGYTAKK-----TYEWMKMLPD----- 130
 QY 65 FARSFYVLMQA---VQPLNISQVPEVDVDDQ-SGLSREIRTLARIEHPLSQDIT 120
 DB 131 -----MSANNVIFDFWELSDQIDLDLTKILP--ELEELAKLPELP---DF- 174
 QY 121 GLEHMLINCSKMLPADITOLNNIPPTQESYVDNPVYKSLVTNCKPYTDKIHKAYKDX 180
 DB 175 -----DKIGENFTFLKSLIFPAEAPGDTPVAAATEAVTATPEASDQFKSSDX 224
 QY 181 NKTRFELMGEELAFKQIRTNVSH--VGQLDIRKPKRFVC-LMDNID-HNHKAQTVK 237
 DB 225 EKV-----DQQLIELRTQKQYKMLERLEKENDAKVYLQDEKGIHQK---IK 273
 QY 238 AVLRDFESMFPISPQF-----LPR 258
 DB 274 KSLIDMYSEVLDLSDFDSNVTQDHLPR 302
 RESULT 3
 EZRA_STRPY STANDARD; PRT; 574 AA.
 ID EZRA_STRPY
 AC 09A059;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Septation ring formation regulator ezra.
 GN EZRA OR SPY0728.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;


```

CC Streptococcus.
OK NCBI_TaxId=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic G., Lyon K.,
  Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Jin S.P.,
  Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
  Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
  Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -!- FUNCTION: Negative regulator of ftsZ ring formation; modulates the
  frequency and position of ftsZ ring formation. Inhibits ftsZ ring
  formation at polar sites. Interacts either with ftsZ or with one
  of its binding partners to promote depolymerization (By
  similarity).
CC -!- SUBCELLULAR LOCATION: Membrane-associated. Colocalized with ftsZ to
  the nascent septal site (By similarity).
CC -!- SIMILARITY: Belongs to the ezra family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
DR EMBL; AB06524; AAC33678.1; -.
DR HAMAP; MF_00728; -; 1.
KW Cell division; Septation; Transmembrane; Coiled coil;
  Complete proteome.
FT DOMAIN 1 7 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 8 26 POTENTIAL.
FT DOMAIN 27 574 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 102 148 Coiled coil (Potential).
FT DOMAIN 276 366 Coiled coil (Potential).
SQ SEQUENCE 574 AA; 66120 MW; 91FF87736186827 CRC64;

Query Match
Best Local Similarity 21.8%; Score 106; DB 1; Length 574;
Matches 61; Conservative 48; Mismatches 113; Indels 58; Gaps 12;

QY 6 SLRYVNIKLSKFGTSRKVPAMPMIRIVMOELQDMFPE---EFDKTSFHKVRHSD 62
DB 126 SIREALNLIKEQ---EEKNSARVTHALD--LYEKLOASISENEDFGSTMPEDIKQMK 179
QY 63 MOPAFSYFYILMSAVOPLNISQVFD--EVDTDQSGVLSDRIRTLATRIHELPLSLQDL- 119
DB 180 IETEFISQFVALNNSGDPVEASVLDRAEHTIALGQITQIPAIYAKLEDDEPPOLDLLE 239
QY 120 TGLEHMLINCSKMLPADITQNLNIPPTQESYYDP-----NLPTVTKSLVTKCKP-VTDK 172
DB 240 TGYRRLL-----EENHYHPEKNIEARFOEIRRESIRANSSELVTLDD 279
QY 173 IHKAYKDKKKYR-----FEIMGEEELIAFKMIRTNVSHVQGLDIRKNPKFYCLNDNT 226
DB 280 IDRAEENHTIQRIDSLYEVEFEREIAAYKVAKKSKMLPYLLEHYKNNQ---LKDEI 336
QY 227 DH-----NKKDAQTKAVLRDPYE---SMFPIPSQDEL 256
DB 337 AELSKYILSETESLTVKAFKDKTEIDSTLAAVQFGL 376

```

```

GN EZRA OR SPVW3_0477 OR SPB1377.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
  Streptococcus.
OK NCBI_TaxId=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Barres S.B., Syva G.L., Barbhan K.D., Lei B., Hoff J.S.,
  Barmes S.B., Syva G.L., Barbhan K.D., Lei B., Hoff J.S.,
  Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
  Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
  phage-encoded toxins, the high-virulence phenotype, and clone
  emergence.";
  Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SGI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
  Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
  Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
  large-scale genomic rearrangement in invasive strains and new insights
  into phage evolution.";
  Genome Res. 13:1042-1055(2003).
CC -!- FUNCTION: Negative regulator of ftsZ ring formation; modulates the
  frequency and position of ftsZ ring formation. Inhibits ftsZ ring
  formation at polar sites. Interacts either with ftsZ or with one
  of its binding partners to promote depolymerization (By
  similarity).
CC -!- SUBCELLULAR LOCATION: Membrane-associated. Colocalized with ftsZ to
  the nascent septal site (By similarity).
CC -!- SIMILARITY: Belongs to the ezra family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
DR EMBL; AE014146; AAM79084.1; -.
DR HAMAP; MF_00728; -; 1.
KW Cell division; Septation; Transmembrane; Coiled coil;
  Complete proteome.
FT DOMAIN 1 7 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 8 26 POTENTIAL.
FT DOMAIN 27 574 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 102 148 Coiled coil (Potential).
FT DOMAIN 276 366 Coiled coil (Potential).
SQ SEQUENCE 574 AA; 66062 MW; 25F87727586825 CRC64;

Query Match
Best Local Similarity 22.5%; Score 104; DB 1; Length 574;
Matches 64; Conservative 48; Mismatches 106; Indels 66; Gaps 14;

QY 6 SLRYVNIKLSKFGTSRKVPAMPMIRIVMOELQDMFPE---EFDKTSFHKVRHSD 62
DB 126 SIREALNLIKEQ---EEKNSARVTHALD--LYEKLOASISENEDFGSTMPEDIKQMK 179
QY 63 MOPAFSYFYILMSAVOPLNISQVFD--EVDTDQSGVLSDRIRTLATRIHELPLSLQDL- 119
DB 180 IETEFISQFVALNNSGDPVEASVLDRAEHTIALGQITQIPAIYAKLEDDEPPOLDLLE 239
QY 120 TGLEHMLINCSKMLPADITQNLNIPPTQESYYDP-----NLPTVTKSLVTKCKP-VTDK 172
DB 240 TGYRRLL-----EENHYHPEKNIEARFOEIRRESIRANSSELVTLDD 279

```

```

RESULT 4
EZRA_STRP3 STANDARD; PRT; 574 AA.
AC 08K39;
DT 15-MAR-2004 (rel. 43, Created)
DT 15-MAR-2004 (rel. 43, Last sequence update)
DT 15-MAR-2004 (rel. 43, Last annotation update)
DE Septation ring formation regulator ezra.

```

QY 173 IHKAYDKKKYR-----FEIMGEELIAFKMIRTN-----VSHV-----COL-DDIRKN 215
DB 280 LDRAREENTHIGERIDSLVEFEREIAAYKAAKSKMLPRYLAHVKNNEQLDEIARL 339
QY 216 PRKFCVCLNDNIDHNHKAQTKAVLRDPYE---SMFPPIPSQPEL 256
DB 340 SRKYTL-----SETESLTVKAFKDKIEXEDSTLVAEQFGL 376

RESULT 5
EZRA_STRPE
ID EZRA_STRPE8 STANDARD; PRT; 574 AA.
AC Q8P1M3;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Separation ring formation regulator ezra.
GN EZRA OR SPYM18_0796.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=186103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.
RT Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
CC -1- FUNCTION: Negative regulator of ftsZ ring formation; modulates the
CC frequency and position of ftsZ ring formation. Inhibits ftsZ ring
CC formation at polar sites. Interacts either with ftsZ or with one
CC of its binding partners to promote depolymerization (by
CC similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Colocalized with ftsZ to
CC the nascent septal site (By similarity).
CC -1- SIMILARITY: Belongs to the ezra family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE010010; AAL97460.1; -
DR HAMAP; MF_00728; -; 1;
KW Cell division; Septation; Transmembrane; Coiled coil;
KW Complete proteome.
FT DOMAIN 1 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 7
FT DOMAIN 8 26
FT DOMAIN 27 574 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 102 148
FT DOMAIN 276 366 Coiled coil (Potential).
SQ SEQUENCE 574 AA; 66061 MW; 76A95C1FC4EB0BF6 CRC64;

Query Match 5.8%; Score 101; DB 1; Length 574;
Best Local Similarity 21.1%; Pred. No. 3.2;
Matches 60; Conservative 48; Mismatches 110; Indels 66; Gaps 12;

DB 6 SURVANKILNSKFGFTSRVPAHMHMDRIYMOQLDMFPE--EFKTFFAKVRHSD 62
DB 126 SIREALNLIKEDQ-----EENKSAKVTHALD--LYEKLQASISENEDNFSTYPEIDKQKN 179
QY 63 MOPASYPYUIMSAAVOPLNISQVDP--EVDTPDSQVLDREIRLTATRIHLPLSLQDL- 119
DB 180 ITPERSQFVALNSSQDPVASEVLDRABEHTIALGQITRQPAIYAKKEDDFPPQDLBLE 239

QY 120 TGEHMLINCSKMLPADITQNLNIPPTOSYDP-----NLPPVTKSLVTKCKP-VTDK 172
DB 240 TGYRRL-----EENYFPKKNIBARQELRESIRASSELVTLTD 279
QY 173 IHKAYDKKKYR-----FEIMGEELIAFKMIRTNVSHVQGLDIRKN----- 215
DB 280 LDRAREENTHIGERIDSLVEFEREIAAYKAAKSKMLPRYLAHVKNNEQLDEIARL 339
QY 216 PRKFCVCLNDNIDHNHKAQTKAVLRDPYE---SMFPPIPSQPEL 256
DB 340 SRKYTL-----SETESLTVKAFKDKIEXEDSTLVAEQFGL 376

RESULT 6
CENE_HUMAN
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC 002234;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis."
RT Nature 359:536-539(1992).
RL ENBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1."
RL J. Cell Biol. 143:49-63(1998).
RN [4]
RP FARNESYLATION.
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RA "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT and CENP-F and alter the association of CENP-E with the
RL microtubules."
RL J. Biol. Chem. 275:30451-30457(2000).
CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -1- SIMILARITY: Belongs to the kinesin-like protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

DR EMBL; Z15005; CAA78727.1; -.

DR PIR; S28261; S28261.

DR HSSP; P17119; 3KAR.

DR Genew; HGNC:1956; CENPE.

DR GK; Q02224; -.

DR MIM; 117143; -.

DR GO; GO:0005699; C.kinetochore; TAS.

DR GO; GO:0005634; C.nucleus; TAS.

DR GO; GO:0008350; F.kinetochore motor activity; TAS.

DR GO; GO:0000067; P.DNA replication and chromosome cycle; TAS.

DR GO; GO:0007079; P.mitotic chromosome movement; TAS.

DR GO; GO:0007080; P.mitotic metaphase plate congression; TAS.

DR InterPro; IPR001752; kinesin_motor.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PR00380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.

DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.

DR Motor protein; Cell division; ATP-binding; coiled coil; Mitosis; Cell cycle; Centromere; Lipoprotein; Prenylation.

KM DOMAIN 1 335 KINESIN-MOTOR.

FT DOMAIN 336 2471 COILED COIL (POTENTIAL).

FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

FT NP BIND 86 93 ARP (BY SIMILARITY).

FT LIPID 2660 2660 S-farnesyl cysteine.

SO SEQUENCE 2663 AA; 312087 MW; CEFCL3860C8C8C8 CR64;

Query Match 5.8%; Score 101; DB 1; Length 2663;
Best Local Similarity 20.5%; Pred. No. 21;
Matches 60; Conservative 47; Mismatches 127; Indels 58; Gaps 10;

QY 33 IIRIMQGEIQDMFPPEPDKTSFHKVRHSEDMQFAFSYFYLYMSAVQPLNISQVDEVDTD 92

DB 1972 IDELKKEIQQLRVKEDVNMHSHKINEMEQLKQEPYLLCKCEMDNFOLTRKQHE--- 2027

QY 93 QSGVLSDEIRTLARIRHLPPLSLDILGLEMLINCSKMLPADITQNLNIPQESYYD 152

DB 2028 -----SLEIRIVAEKREDELRIKESLSKMERQDFATLRKEMLARQNHQVKEKRLISD 2082

QY 153 ENLPVYTSGLVTNCGPVTDKIHKAYKDKNKY----REFINGEEELAFKMTRTNVSHVG 207

DB 2083 -GQCHLMESLRKESCRKIKEL-KRYSEMDHYECNLRLSLDEKIEFHRIMKKLKLYLS 2140

QY 208 QDDDIRKNRKFVNCND-----NIDNNHDAQV----- 235

DB 2141 VYTKIKEQHE--CLNKFEKMPIDVEVEKOKELLIKIQHLOQDCDVPSELRDLKLNQMD 2198

QY 236 --VKAFLRDVYESNFP-IPSGELPFEYRNREFLHNLQEMRPAVY---DKLK 281

DB 2199 LHIEELIKDFSEFPSTIKTEFQ-QVLSNRKEMQGLEMLNTRFDIEKXK 2248

RESULT 7

TATR_NPYBM STANDARD; PRT; 587 AA.

AC P33245;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Trans-activating transcriptional regulatory protein (Immediate early protein 1) (IE-1).

GN IE1.

OS Bombyx mori nuclear polyhedrosis virus (BmNPV).

CC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;

CC Nucleopolyhedrovirus.

OX NCBI_Taxid=10458;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92162753; PubMed=1536885;

RA Huybrechts R., Guarino L., van Brusseel M., Vuleteke V.;

RT "Nucleotide sequence of a transactivating Bombyx mori nuclear

RT polyhedrosis virus immediate early gene.";

RL Biochim. Biophys. Acta 1129:328-330(1992).

CC -1- FUNCTION: Regulatory transcriptional protein, which trans-activates gene expression from early baculovirus promoters. Can also trans-activate its own promoter, suggesting that it is autoregulated during normal infection of insect cells.

CC -1- SIMILARITY: TO OTHER BACULOVIRUSES IE-1.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

DR EMBL; X58442; CAA1348.1; -.

DR PIR; S20596; S20596.

DR InterPro; IPR005092; TATR.

DR Pfam; PF03430; TATR; 1.

KW Early protein; Transcription regulation; Trans-acting factor.

SO SEQUENCE 587 AA; 67495 MW; 5896B0C9625ACB59 CR64;

Query Match 5.7%; Score 99.5; DB 1; Length 587;
Best Local Similarity 23.4%; Pred. No. 4.2; Indels 53; Gaps 11;
Matches 58; Conservative 34; Mismatches 103; Indels 53; Gaps 11;

QY 39 QELQDMFPPEE-----FDKTSFHKVRHSEDM-QFAFSYFYLYMSAVQPLNI-----SQ 84

DB 189 QEITHYFTNDFARYLMRFDNDYNSNRFSQHMSETGYFVVKSEVKEPPEILFAKYGVN 248

QY 85 VPDEVDTQSGVLSDEIRTLARIRH-----ELPISLDILGLEMLINCSK 131

DB 249 VVVEY-TNNYVMQDNRFVVTFDKIRFMISYNLKETGELIIPHS-ODVCNDETPAQCCKR 306

QY 132 MLPADITQNLNIPPTQESYYDPLPPTKSLVT-----NCKPVTDKIHKAYKDKNKY 183

DB 307 CHEVDVA--HTFPAALTSVFNLDMYAAQTFTVLLSLGSRKCGFLSLKLYEYQDKNLF 364

QY 184 RPEIM-----GEEELAFK-MRTNVSHVGGQLDIRK-----NPRFVCLNIDRN 229

DB 365 TLPIMSRKESNEIEFASNNFVSPVYSQILKXSESIRKVKPDPNPXYVDNLILVN 424

QY 230 HKDAQTVK 237

DB 425 KKSTLYRK 432

RESULT 8

GBA2 SOYBN STANDARD; PRT; 385 AA.

AC P93163;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Guanine nucleotide-binding protein alpha-2 subunit (GP-alpha-2).

GN GPA2 OR GA2.

OS Glycine max (Soybean).

CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; OC Eustrodi; I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

OX NCBI_Taxid=3847;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Corsoy;

RX MEDLINE=97156166; PubMed=9002626;

RA Gecor C., Lam E., Cejudo F.J., Romero L.C.;

RT "Isolation and analysis of the soybean GBA2 gene (cDNA), encoding a new member of the plant G-protein family of signal transducers.";

RT Plant Mol. Biol. 32:1227-1234(1996).

CC -1- FUNCTION: Guanine nucleotide-binding proteins (G proteins) are involved as modulators or transducers in various transmembrane signaling systems.

```

CC -1- SUBUNIT: G proteins are composed of 3 units; alpha, beta and
CC gamma. The alpha chain contains the guanine nucleotide binding
CC site.
CC -1- SIMILARITY: Belongs to the G-alpha family. Subfamily 2 (G(s)).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X55882; CAA64934.1; -.
CC HSPSP; P10824; IAS3.
CC InterPro: IPR001019; Gprotein_alpha.
CC Pfam: PF00503; G-alpha; 1.
CC PRINTS: PR00318; GPROTEIN.
CC ProDom: PD000281; Gprotein_alpha; 1.
CC SMART: SM00275; G_alpha; 1.
CC GTP-binding; Transducer; ADP-ribosylation.
CC NP_BIND 47 54 GTP (BY SIMILARITY).
CC FT NP_BIND 220 224 GTP (BY SIMILARITY).
CC FT NP_BIND 289 292 GTP (BY SIMILARITY).
CC FT MOD_RES 192 192 ADP-RIBOSYL[1] (BY ACTION OF CTX).
CC SEQUENCE 385 AA; 45096 MW; 7721B1B307067075 CRC64;

Query Match 5.7%; Score 99; DB 1; Length 385;
Best Local Similarity 22.6%; Pred. No. 2.7;
Matches 57; Conservative 44; Mismatches 101; Indels 50; Gaps 11;

QY 37 VNOELQDMPEEFEDDT--SFHVSSEDMQFAFYFYLLSAVOPINI---GQVPEV 89
DB 56 IRKQIKLFGTGFDAELKSYPIVHA-----NVYOTITLHDSKEFRQN 101
QY 90 DTDQSGVLSDRIRLRIHPLSLQDTGLEHMLINCSKMLPADITQINIPROES 149
DB 102 DVDSKRYVINSNKXIGKLEIGRL-DYPL-----SKELAQEIENMKPDAIOET 153
QY 150 Y---YDPMLEPVTKSLVNTCKEPTVKIKHAYKDKKY-RPEIMGEELAFPMIRTN-VSH 204
DB 154 YARGSELDQIDCTDYFENLQRLSDANVPTEDVLYARVTTGVEIQSFVGENKSD 213
QY 205 VVGQDDI---RKNPRKFCVCLNDNIDHNKDAQTVKAVLRDPFESMPISQFELPREYR 261
DB 214 EYVRLPFDVGGQNRERRKMIHLPEGV---SAVIFCAIASEVDQTLF-----EDENR 260
QY 262 NRELMHHEIQEW 273
DB 261 NRMETKELFEW 272

RESULT 9
MYH9_RAT STANDARD; PRT; 1961 AA.
AC Q62812;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Cellular myosin appears to play a role in cytokinesis,
CC cell shape, and specialized functions such as secretion and
CC capping.

```

```

CC -1- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
CC chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
CC regulatory light chain subunits (MLC-2).
CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U1463; AA474950.1; -.
CC HSPSP; P10587; 1BR2.
CC InterPro: IPR000048; IQ_region.
CC InterPro: IPR001609; myosin_head.
CC InterPro: IPR004009; Myosin_N.
CC InterPro: IPR002928; Myosin_tail.
CC Pfam: PF00612; IQ; 1.
CC Pfam: PF00663; myosin_head; 1.
CC Pfam: PF02736; Myosin_N; 1.
CC Pfam: PF01576; Myosin_tail; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC ProDom: PD000355; myosin_head; 1.
CC SMART: SM00015; IQ; 1.
CC SMART: SM00242; MYSC; 1.
CC PROSITE: PS50096; IQ; 1.
CC KX Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
CC Coiled coil; Multigene family.
CC DOMAIN 1 778 MYOSIN HEAD-LIKE.
CC FT DOMAIN 779 808 IQ.
CC FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
CC FT NP_BIND 174 181 ATP (POTENTIAL).
CC FT DOMAIN 654 676 ACTIN-BINDING.
CC SEQUENCE 1961 AA; 226336 MW; 9B9876D9681FB19E CRC64;

Query Match 5.6%; Score 97; DB 1; Length 1961;
Best Local Similarity 28.2%; Pred. No. 29;
Matches 35; Conservative 26; Mismatches 51; Indels 12; Gaps 5;

QY 102 IRT-LATRIHEPLSLQDTGLEHMLINCSKMLPADITQL-NNIPPOSSYYDPNLPVT 159
DB 1267 VRTLELDKSKIQVELDSTGLNQSXSXTDFSLBSQLODTGLQENRQKL- 1325
QY 160 KSLVNTCKEPTVKIKHAYKDKKYRPEIMGEELAFPMIRTNVSHVGGQDDIRKPKKF 219
DB 1326 -SLSTYKQMED-----EKNSFREOLEEERBAKRNLEKQIATLHAQVTDMKKMEDG 1377
QY 220 V-CL 222
DB 1378 VGCL 1381

RESULT 10
VAC1_YEAST STANDARD; PRT; 515 AA.
AC P32659;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE VAC1 protein.
GN VAC1 OR PEP7 OR VPS19 OR YDR323C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=LW148;
RA MEDLINE=92112720; PubMed=1730622;
RX Weisman J.S., Wickner W.;
RT "Molecular characterization of VAC1, a gene required for vacuole
RT inheritance and vacuole protein sorting.";
RL J. Biol. Chem. 267:618-623(1992).
CC -1- FUNCTION: REQUIRED FOR VACUOLE SEGREGATION AND VACUOLE PROTEIN
CC SORTING. POSSIBLY PART OF A COMPLEX WHICH TETHERS THE VACUOLE
CC MEMBRANE TO MICROTUBULES, EITHER DIRECTLY OR VIA KINESIN OR
CC DYNEIN-LIKE MOTOR PROTEINS. PROBABLY FUNCTIONS IN SEVERAL
CC INTERORGANELLE TRAFFIC PATHWAYS.
CC -1- SUBCELLULAR LOCATION: PERIPHERALLY BOUND ELEMENT OF THE
CC CYTOSKELETON OR VACUOLE (POTENTIAL).
CC -1- SIMILARITY: contains 1 C2H2-type zinc finger.
CC -1- SIMILARITY: contains 2 FYVE-type zinc fingers.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-109 IS THE
CC INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80596; AAA35203.1; -.
DR GEMOnline; 140815; -.
DR SGD; S0002731; PEP7.
DR GO; GO:0005737; Cytoplasm; IDA.
DR GO; GO:0019897; Cytoplasmic to plasma membrane; IDA.
DR GO; GO:0006896; P:Golgi to vacuole transport; IGI.
DR GO; GO:0006904; P:nonspecific vesicle docking; IMP.
DR GO; GO:0006906; P:nonspecific vesicle fusion; IMP.
DR InterPro; IPR007087; Znf_FYVE.
DR InterPro; IPR00306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR Pfam; PF0096; ZF_C2H2; 1.
DR SMART; SM00064; FYVE; 2.
DR SMART; SM00355; ZNF_C2H2; 1.
DR PROSITE; PS01178; ZF_FYVE; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR Zinc-finger; Repeat.
DR ZNF_FING 6 29 C2H2-TYPE.
FT ZN_FING 72 137 FYVE-TYPE 1.
FT ZN_FING 215 297 FYVE-TYPE 2.
SQ SEQUENCE 515 AA; 59469 MW; 02C204E1BCA8ACC CRC64;
Query Match 5.5%; Score 96; DB 1; Length 515;
Best Local Similarity 22.4%; Pred. No. 6;
Matches 57; Conservative 39; Mismatches 96; Indels 62; Gaps 12;
QY 28 HYP-----HVIDRIVMGLDMFPEEPDKTSFKVHSHSEDMQAFSYFYLYMSAVPL 80
DB 281 HPIPIRCLSHCIDMW-----FIGKFKK-----DVKNPLSGITAKTDSVO-- 321
QY 81 NISQVFEVDYDQSGVLSDRIRTLATRIHEL-PUSLODLTGLBMLINCSKMLPADITQ 139
DB 322 NISKYIDSLPIFEDSLNSIKVETAKDSENTLDPRGNLMDLARLRKYLINSFNLVNTLTRQ 381
QY 140 L-----NNIPRQESYVDNLPVPTKSLVNTCKCPYDKTHKXKDKKXKRF 185
DB 382 LLSVEPQSHLERQONSIKIASAAYINEXILPL-KSLPAIINP--EGHKXNEGOQKXEP 437
QY 186 EI-----MGEERIAFK-----MIRTNVSHV-GQLDDIRKNFR--KEVCINDNDH 228
DB 438 EYKXLSQMLTENLTLTKVXKLEBEELMWLKEQSYLLESTIOPYKKQRRLBEIVTLKNKEE 497
QY 229 HKKDAQTYKAYLRD 242
DB 498 LHSRIHTVOSKLG 511

```

```

RESULT 11
RBP2_PLAVB STANDARD; PRT; 2867 AA.
AC 000759; Q9NZM3;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Reticulocyte binding protein 2 precursor (PVRBP-2).
GN RBP-2 OR RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=12273;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.
RX MEDLINE=20299192; PubMed=10838229;
RA Galinski M.R., Xu M., Barnwell J.W.;
RT "Plasmodium vivax reticulocyte binding protein-2 (PVRBP-2) shares
RT structural features with PVRBP-1 and the Plasmodium yoelii 235 kDa
RT thoptry protein family.";
RL Mol. Biochem. Parasitol. 108:257-262(2000).
RN [2]
RP SEQUENCE OF 1189-2439 FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites.";
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to
CC human reticulocyte cells.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF184623; AAF76525.1; -.
DR HSP; P03069; ICKM.
KW Malaria; Receptor; Signal; Transmembrane; Repeat.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 2867 RETICULOCYTE BINDING PROTEIN 2.
FT DOMAIN 22 2805 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2806 2826 POTENTIAL.
FT DOMAIN 2827 2867 CYTOSOLASMIC (POTENTIAL).
FT DOMAIN 44 133 ASN-RICH.
FT DOMAIN 560 758 LYS-RICH.
FT DOMAIN 1112 1285 LYS-RICH.
FT DOMAIN 2758 2785 7 X 4 AA TANDEM REPEATS OF H-D-D-T.
FT REPEAT 2758 2761 1.
FT REPEAT 2762 2765 2.
FT REPEAT 2766 2769 3.
FT REPEAT 2770 2773 4.
FT REPEAT 2774 2777 5.
FT REPEAT 2778 2781 6.
FT REPEAT 2782 2785 7.
SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFFD3 CRC64;
Query Match 5.5%; Score 95; DB 1; Length 2867;
Best Local Similarity 20.5%; Pred. No. 66;
Matches 56; Conservative 55; Mismatches 78; Indels 84; Gaps 14;
QY 7 LRYVNK-----INSKFGFTSKVPAFPMHIDRIVMQELDMFPEEPDKTSFKVYR 59
DB 2193 LAYIKKNYEDTVQDVLINHEHFTKQVSNH-----PNNFDKSN-----K 2232
QY 59 HSEDMQAFSYFYLYMSAVOPLNISQVFD-EVDYDQSGVLSDRIRTLATRIHEL-PUSL 116
DB 2233 SSEELTKAVYDSKTIISKLGV-IIEVNNTENNTIES--SAKETLEALYNELKNNKTS 2288
QY 117 QDL-----TGLEHMLINCSKMLPADITQL-NNIPRQESYVDNLPVPTKSLVNTCKRV 169

```



```

FT HELIX 80 82
FT TURN 83 84
FT STRAND 87 87
FT TURN 88 89
FT HELIX 96 108
FT TURN 109 110
FT STRAND 113 115
FT TURN 120 123
FT STRAND 130 131
FT TURN 134 139
FT HELIX 140 142
FT TURN 145 147
FT HELIX 152 166
FT TURN 167 167
FT STRAND 170 175
FT TURN 178 179
FT HELIX 182 197
FT HELIX 218 223
FT TURN 224 225
FT HELIX 226 234
FT STRAND 235 236
FT TURN 242 243
FT STRAND 244 245
FT STRAND 248 255
FT TURN 257 258
FT STRAND 261 269
FT HELIX 273 276
FT TURN 277 277
FT TURN 281 282
FT STRAND 286 286
FT HELIX 287 295
FT HELIX 298 302
FT TURN 309 311
FT TURN 313 314
FT TURN 316 317
FT HELIX 327 341
FT TURN 342 342
FT HELIX 345 363
FT STRAND 367 367
FT STRAND 376 376
FT TURN 381 390
FT TURN 391 391
FT HELIX 394 402
FT HELIX 419 449
FT STRAND 458 464
FT STRAND 474 474
FT HELIX 476 506
FT TURN 507 508
FT HELIX 521 528
FT HELIX 536 543
FT HELIX 551 561
FT TURN 562 564
FT STRAND 566 570
FT STRAND 579 584
FT TURN 585 585
FT STRAND 587 592
FT TURN 594 595
FT HELIX 596 601
FT HELIX 606 613
FT TURN 614 614
FT HELIX 618 623
FT TURN 624 624
FT HELIX 658 674
FT TURN 675 675
FT STRAND 677 684
FT TURN 692 693
FT HELIX 697 706
FT TURN 707 708
FT HELIX 709 718
FT STRAND 722 725
FT HELIX 726 733

```

```

FT HELIX 734 737
FT TURN 738 739
FT HELIX 748 759
FT TURN 760 760
FT HELIX 763 765
FT STRAND 766 768
FT STRAND 772 775
FT TURN 777 778
FT HELIX 779 787
SQ SEQUENCE 1978 AA; 228663 MW; B7B6C923E5273D93 CRC64;

```

```

Query Match 5.4%; Score 94; DB 1; Length 1978;
Best Local Similarity 23.5%; Pred. No. 50;
Matches 42; Conservative 41; Mismatches 72; Indels 24; Gaps 8;

```

```

QY 81 NISQVDEVD-----TDQGVLSPRE-IRT-LATRIHELPLSLDPLGLEMLIN 128
DB 1247 SLSQAKQDVHKKKLEVLQDLQSKVSDGERVTELNKVKLQIENVNVLSLNEMES 1306
QY 129 CSMPLPADITOL-NIIPFOESYDPNLPVTKSLVTNCKPYTDKIHAKYDKKXRFET 187
DB 1307 KNIKLTQDVATLGSQLODTQEL-----LOEETROKLN---VTYKLRQLEDPKSLQ-EQ 1356
QY 188 MGEETAFKMRITNVSHVVGOLDIRKRNKRFVCLNDNIDHVKDAQ-TVKAVLDFVE 245
DB 1357 LDEEVAKQKNERHISTITQLSDSKKLQEFATVETMEBKGLQREIESLTOQFEH 1415

```

RESULT 15

```

SYN_BORBU STANDARD; PRT; 353 AA.
ID OS1038:
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan-tRNA ligase)
DE (TrpRS).
GN TRPS OR TRSA OR BR0005.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxId=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Suton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gunn M.,
RA Peterson J., Tomb J.-F., Fleischmann R.D., Richardson D., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman C.,
RA Uterback T., Matthey L., McDonald L., Artlich P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RT Nature 390:580-586 (1997).
RL

```

```

-1- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +
diphosphate + L-tryptophanyl-tRNA(Trp).
-1- SUBUNIT: Homodimer (By similarity).
-1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; AE001115; AAC66398.1; -.
CC DR PIR; E70100; E70100.
CC DR HSSP; P00953; 1D2R.

```


This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:05:03 ; Search time 21.4324 Seconds

(without alignment)
4828,666 Million cell updates/sec

Title: US-10-657-280-2

Perfect score: 1731

Sequence: 1 DTFADSLRYVKKILNSKRGF.....RKIPRRRIKESAPNIRV 328

Scoring table: BLOSUM62

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1731	100.0	490	4	Q86TQ2	Q86TQ2 homo sapien
2	1731	100.0	950	4	Q9ULJ2	Q9ULJ2 homo sapien
3	1290	74.5	248	4	Q9NPW9	Q9NPW9 homo sapien
4	704	40.7	132	4	Q9NV34	Q9NV34 homo sapien
5	581	33.6	184	11	Q61340	Q61340 mus musculu
6	506	29.2	666	5	Q8V553	Q8V553 drosophila
7	506	29.2	666	5	Q8V553	Q8V553 drosophila
8	113.5	6.6	545	2	Q68215	Q68215 neisseria m
9	113.5	6.6	545	2	Q84D00	Q84D00 neisseria m
10	113.5	6.6	545	2	Q84C29	Q84C29 neisseria m
11	113.5	6.6	545	2	Q83U59	Q83U59 neisseria m
12	113.5	6.6	545	16	Q9JW88	Q9JW88 neisseria m
13	111	6.4	1048	5	Q26023	Q26023 plasmodium
14	111	6.4	3119	5	Q25857	Q25857 plasmodium
15	111	6.4	3119	5	Q814T3	Q814T3 plasmodium
16	106.5	6.2	483	5	Q86KMO	Q86KMO dictyosteli

17	106	6.1	574	16	Q9A0K9	Q9A0K9 streptococc
18	106	6.1	739	8	Q9M079	Q9M079 physarum po
19	106	6.1	2148	5	Q9BUD3	Q9BUD3 physarum po
20	105	6.0	666	13	Q90XS4	Q90XS4 xenopus lae
21	104.5	6.0	1999	11	Q63731	Q63731 rattus norv
22	104	6.0	424	5	Q81188	Q81188 streptococc
23	104	6.0	574	16	Q8K839	Q8K839 streptococc
24	103	6.0	621	16	Q813R0	Q813R0 bacillus ce
25	103	6.0	690	17	Q57915	Q57915 pyrococcus
26	102.5	5.9	676	5	Q7YX23	Q7YX23 cryptospori
27	102.5	5.9	706	8	Q98894	Q98894 guillardia
28	102.5	5.9	515	3	Q99229	Q99229 saccharomy
29	101.5	5.9	301	16	Q8Y9G9	Q8Y9G9 listeria mo
30	101.5	5.9	408	12	Q91575	Q91575 fowlpox vir
31	101	5.8	375	10	Q7XW89	Q7XW89 oryza sativ
32	100.5	5.8	574	16	Q8B1M3	Q8B1M3 streptococc
33	100.5	5.8	265	16	Q8B241	Q8B241 wigleswort
34	100.5	5.8	468	5	Q86BM6	Q86BM6 drosophila
35	100	5.8	5987	5	Q81DN3	Q81DN3 plasmodium
36	99.5	5.7	983	4	Q9H0L3	Q9H0L3 homo sapien
37	99	5.7	723	16	Q8RH48	Q8RH48 fusobacteri
38	98	5.7	5687	5	Q81UH4	Q81UH4 plasmodium
39	97.5	5.6	735	16	Q51543	Q51543 borrelia bu
40	97.5	5.6	893	4	Q81YMO	Q81YMO homo sapien
41	97.5	5.6	1498	16	Q92FP9	Q92FP9 listeria in
42	97	5.6	568	5	Q18679	Q18679 caenorhabdi
43	97	5.6	620	16	Q99227	Q99227 streptococc
44	97	5.6	776	16	Q8B816	Q8B816 shewanella
45	97	5.6	942	16	Q91114	Q91114 streptomyc

ALIGNMENTS

RESULT 1	ALIGNMENTS
Q86TQ2	PRELIMINARY; PRT; 490 AA.
AC Q86TQ2	
DT 01-JUN-2003 (T-EMBLrel. 24, Created)	
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)	
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)	
DE Similar to v-mat musculoepitrotoic fibrosarcoma oncogene family,	
DE protein B (Avian) (Fragment).	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=Liver;	
RA Strauberg R.;	
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.	
DR EMBL: BC042615; AA042615.1; -	
DR GO: GO:0005509; F-actin ion binding; IEA.	
DR InterPro: IPR002046; EF-hand.	
DR Pfam: PF00036; efhand. 1.	
DR PROSITE: PS00018; EF_HAND; 1.	
FT NON TER	
FT SEQUENCE 490 AA; 57412 MW; 37B19FE0D1259AD2 CRC64;	
Query Match	100.0%; Score 1731; DB 4; Length 490;
Best Local Similarity	100.0%; Pred. No. 1e-128;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 DTFADSLRYVKKILNSKRGFTSRKVPAMHPHMDIRIVMOELQMPFEEFDKTSFHKVRSS 60	
DB 163 DTFADSLRYVKKILNSKRGFTSRKVPAMHPHMDIRIVMOELQMPFEEFDKTSFHKVRSS 222	
QY 61 EDMQAFASFYRYMSAVQPNISQVDEVDTDOSGVLSDREIRTTLTRTHEPLSLQDLT 120	
DB 223 EDMQAFASFYRYMSAVQPNISQVDEVDTDOSGVLSDREIRTTLTRTHEPLSLQDLT 282	
QY 121 GLBHLINCSKMLPADITQANNIPQOESYYDBNLDPVTKSLVTNCKPVTDKIKKAYKDK 180	

```

Db      283 GLEHMLINCSKMLPADITQLNNI PPTQESYVDNIPVTKSLVTNCKEPTDKIHKAYKDK 342
Qy      181 NKCPREIMGEELAFPMIRTNVSHVVGQDDIRKPKRFVCLNDINDHNHKAQTVKAVL 240
Db      343 NKTRFIMGEELAFMIRTNVSHVVGQDDIRKPKRFVCLNDINDHNHKAQTVKAVL 402
Qy      241 RDPFESMPFIPSOFLPREYRNFPLMHLEQEMRAYRDKLKFVTHCVLATLIMFTIFSFF 300
Db      403 RDPFESMPFIPSOFLPREYRNFPLMHLEQEMRAYRDKLKFVTHCVLATLIMFTIFSFF 462
Qy      301 AEOLIAKKRIFPRRIHKEASPNRIRV 328
Db      463 AEOLIAKKRIFPRRIHKEASPNRIRV 490

RESULT 2
Q9ULI2 PRELIMINARY; PRT; 950 AA.
AC Q9ULI2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA1208 (Fragment).
GN KIAA1208.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:337-345 (1999).
DR EMBL; AB033034; BAA86522.2; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00036; efhand; 1.
DR Pfam; PF00066; notch; 1.
DR SMART; SM0004; NT; 2.
DR PROSITE; PS00018; EF_HAND; 1.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 950 AA; 108785 MM; F653BDFAEAC7503 CRC64;

Query Match
Best Local Similarity 100.0%; Score 1731; DB 4; Length 950;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DTFADSLRYVKNILNSKRGFTSRKVPAMHPMIDRIVMQELQDVPPEFDKTSFKVYRHS 60
Db      623 DTFADSLRYVKNILNSKRGFTSRKVPAMHPMIDRIVMQELQDVPPEFDKTSFKVYRHS 682
Qy      61 EDMOPAFSYFYLYMSAVOPLNISQVFDVDTDQSGVSDREIRLATIHELPLSIDLT 120
Db      683 EDMOPAFSYFYLYMSAVOPLNISQVFDVDTDQSGVSDREIRLATIHELPLSIDLT 742
Qy      121 GLEHMLINCSKMLPADITQLNNI PPTQESYVDNIPVTKSLVTNCKEPTDKIHKAYKDK 180
Db      743 GLEHMLINCSKMLPADITQLNNI PPTQESYVDNIPVTKSLVTNCKEPTDKIHKAYKDK 802
Qy      181 NKCPREIMGEELAFPMIRTNVSHVVGQDDIRKPKRFVCLNDINDHNHKAQTVKAVL 240
Db      803 NKCPREIMGEELAFPMIRTNVSHVVGQDDIRKPKRFVCLNDINDHNHKAQTVKAVL 862
Qy      241 RDPFESMPFIPSOFLPREYRNFPLMHLEQEMRAYRDKLKFVTHCVLATLIMFTIFSFF 300

```

```

Db      863 RDPFESMPFIPSOFLPREYRNFPLMHLEQEMRAYRDKLKFVTHCVLATLIMFTIFSFF 922
Qy      301 AEOLIAKKRIFPRRIHKEASPNRIRV 328
Db      923 AEOLIAKKRIFPRRIHKEASPNRIRV 950

RESULT 3
Q9NPM9 PRELIMINARY; PRT; 248 AA.
AC Q9NPM9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP762B26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RA Bloembergen H., Beecher M., Brandt P., Mewes H.W., Weill B., Wiemann S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DDSI databases.
DR EMBL; AL535588; CAB94874.1; -.
DR PIR; T50618; T50618.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; 1.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 248 AA; 29336 MM; D4C4CA6C4505522 CRC64;

Query Match
Best Local Similarity 74.5%; Score 1290; DB 4; Length 248;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      85 VFPEVDTDQSGVSDREIRLATIHELPLSIDLTGLEHMLINCSKMLPADITQLNIP 144
Db      5 VFPEVDTDQSGVSDREIRLATIHELPLSIDLTGLEHMLINCSKMLPADITQLNIP 64
Qy      145 PPTQESYVDNIPVTKSLVTNCKEPTDKIHKAYKDKNKYRPFIMGEELAFMIRTNVSH 204
Db      65 PPTQESYVDNIPVTKSLVTNCKEPTDKIHKAYKDKNKYRPFIMGEELAFMIRTNVSH 124
Qy      205 VVGQDDIRKPKRFVCLNDINDHNHKAQTVKAVLRDPFESMPFIPSOFLPREYRNF 264
Db      125 VVGQDDIRKPKRFVCLNDINDHNHKAQTVKAVLRDPFESMPFIPSOFLPREYRNF 184
Qy      265 LHMHELOEMRAYRDKLKFVTHCVLATLIMFTIFSFAEOLIAKKRIFPRRIHKEASPN 324
Db      185 LHMHELOEMRAYRDKLKFVTHCVLATLIMFTIFSFAEOLIAKKRIFPRRIHKEASPN 244
Qy      325 RIRV 328
Db      245 RIRV 248

RESULT 4
Q9NV34 PRELIMINARY; PRT; 132 AA.
AC Q9NV34;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ10959.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```
RP SEQUENCE FROM N.A.
RC Tissue=Placenta;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuko Y., Oshima A.,
RT "NEO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001821; EAA91926.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 132 AA; 16083 MW; 92B5DB4E062AD07 CRC64;

Query Match 40.7%; Score 704; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 3,66-48;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 MIRNVSHVVGQLDIRKPKFVCCLNDHNDHQAQTVXVLRDPFESMFPISQPEL 256
DB 1 MIRNVSHVVGQLDIRKPKFVCCLNDHNDHQAQTVXVLRDPFESMFPISQPEL 60
QY 257 PREVENRFLHMEHLEOMRAYRDKLKEWTHCVATLIMFTISFFAEOLIALKRIKIFPERR 316
DB 61 PREIKRFLHMELEOMRAYRDKLKEWTHCVATLIMFTISFFAEOLIALKRIKIFPERR 120
QY 317 IHKEASPNRIRV 328
DB 121 IHKEASPNRIRV 132

RESULT 5
Q61340 PRELIMINARY; PRT; 384 AA.
AC Q61340;
ID Q61340;
AC 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Basic domain/leucine zipper transcription factor (fragment).
GN MAFB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95094266; PubMed=8001130;
RA Cordes S.P., Barsh G.S.;
RT "The mouse segmentation gene kr encodes a novel basic domain-leucine
RT zipper transcription factor.";
RL Cell 79:1025-1034(1994).
DR EMBL; U36434; AAA65688.1; -.
DR PIR; I49528; I49528.
DR MGD; MGI:104555; MafB.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:0003579; P:regulation of transcription from Pol II pro. .; IDA.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 1.
DR PROSITE; PS00018; EF_HAND; 1.
FT NON_TER
SQ
SEQUENCE 384 AA; 43754 MW; 51F473C8807A7E55 CRC64;

Query Match 33.6%; Score 581; DB 11; Length 384;
Best Local Similarity 97.4%; Pred. No. 7,36-38;
Matches 112; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
RESULT 6
Q085X14 PRELIMINARY; PRT; 666 AA.
ID Q085X14
AC Q085X14;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE R85033P.
GN CG8027.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STPAIR=Berkely.
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
RA Champ M., Chavez C., Dorsett V., Dreesen D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guatin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.U., Nuno U., Paclio U., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceuliker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RU EMBL; AY089618; AL90356.1; -.
DR FLYBase; FBgn0033392; CG8027.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00066; notch; 1.
SQ
SEQUENCE 666 AA; 77745 MW; F6FDB6D1C1C39248 CRC64;

Query Match 29.2%; Score 506; DB 5; Length 666;
Best Local Similarity 33.6%; Pred. No. 1,26-31;
Matches 108; Conservative 70; Mismatches 133; Indels 10; Gaps 5;

QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAMPHVIDRIWQELQDMPEEFDKTSFHKVRHS 60
DB 348 DIYSHSLIATNMILNAYGKARVLAHYGLIDKQIVEMQRRFQQLDTHQFFRP 407
QY 61 EDMQFAFSFYFLMSAVQPLNISQVDEVDTDGSGVLSDEIRFTLATRIHEPLSLQDLT 120
DB 408 TDLQYAFAYVSLMSSTKWSVEIRFDESDTGSATWSDEVRFTFLRIYQPLDWSAMR 467
QY 121 GLEHMLNCSK---MLPADITQLNNIPPTQESYDPNLPPVYKSLVTNCKPVTDKIHA 176
DB 468 YFEFVQNCNTRNLGMHLKVDVVEHSTL--VYETYESNLPITRIIDLVAQPLLAELAAAN 525
QY 177 YKDKNRYFEIIGE--EELAKKIRNVSHVVGQLDIRKPKFVCLNDNDIDNEF-KDA 233
DB 526 FAVRPKKNFVSPKRSKSHSNFEMLTSLTEVVESLRLRRNPKFKKINCINDANGEEN 585
QY 234 QYKAVLRDYEEMFPIPSQFELPRRYRPFILMHGELQEMRAYRDKLKEWTHCVATLIM 293
DB 586 EMYRHLHLDYLSLFFRRRSKFEELPQYRNRFBSWRDFOVKR-RKAAVLVIGVSLLV 644
QY 294 FTFSFPAEQILALKRIKIFPR 314
DB 645 VCLLRFCVHKAKLVRCVQR 665

RESULT 7
Q09V553 PRELIMINARY; PRT; 666 AA.
ID Q09V553
AC Q09V553;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE CG8027-PA.
GN CG8027.
OS Drosophila melanogaster (fruit fly).
```

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratunga C., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Abri J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Bencze P.V., Bernier B.P., Bhandari D., Bolshakov S.,
 RA Borokov D., Botchan M.R., Bouck J., Brockstein P., Brodtier P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong P., Gerreil J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodra C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mody C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Sidenfarms I., Simpson M., Skusek M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.T., Wasserman D.A., Weinstein G.W., Weissbach J.,
 RA Williams S.M., Woodgett M., Woodgett K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochownik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
 RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
 RA Harris N.L., Krommiller B., Marshall B., Milburn G.H., Richter J.,
 RA Russo S., Seattle S.M.J., Smith E., Shu S., Smutnicka F.,
 RA Whitefield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
 RA Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Flybase;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Flybase;
 RL Submitted (JAN-2003) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AAC00834; AAF58987.2;
 DR GO, GO:0016020; C:membrane; IEA;
 DR GO, GO:0030154; P:cell differentiation; IEA;
 DR InterPro, IPR000800; Notch_dom.
 DR Pfam, PF00066; notch; 1.
 DR SMART, SM00004; NL; 1.
 SQ SEQUENCE 666 AA; 77731 MW; EA232E5C7C754FF6 CRC64;

Query Match 29.2%; Score 506; DB 5; Length 666;
 Best Local Similarity 33.6%; Pred. No. 1.2e-31;
 Matches 108; Conservative 70; Mismatches 133; Indels 10; Gaps 5;
 QY 1 DTFADSLRYVKNKLSKFGFTSRKVPANPMIDRIWQELQDMPPEEFKTSFEKVRHS 60
 DB 348 DIYSHLLIATNMLNINAYGFKARHVAHAGFLDKDQIVBAMORRFRQCLIDPAHOFRA 407
 QY 61 EDMQFAFSFVYVYMSAVQPLNTSQVPEVDVDTQSGVLSREIRLTATRIHELPLSQDIT 120
 DB 408 TDQVAFVY 467
 QY 121 GLEHMLINCSK---MLPADITQLNIPPTOSYVDPNLPVTKSLVNTCKEPTDKIKHA 176
 DB 468 YFEBAVQVQNRNLGMELKVDYVESHVL--VYREYEDSNLPITRDLVYVCPILAEBALAN 525
 QY 177 YKQAKVY 233
 DB 526 FAVRPFYVY 585
 QY 234 QTVKAVLPDYESEMPDIPQPELPRYRNFPHMELOEMRAYRDKLKFTGCVLATLIM 293
 DB 586 EMYRHLLEDYLSFFPRRREKTFELPQYRRRPFESMDFQMKR-RKRAVLVIGVSYLLV 644
 QY 294 FTTFSPFAEQILAKKXIPR 314
 DB 645 VCLRFMCHHKAKLVRCVQR 665
 RESULT 8
 ID 068215 PRELIMINARY; PRT; 545 AA.
 AC 068215;
 DT 01-AUG-1998 (TEMBLrel. 07, Created)
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE SAGB.
 GN *Neisseria meningitidis* (serogroup A).
 OS *Neisseria meningitidis* (serogroup A).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 ON NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=F8229;
 RX MEDLINE=98175678; PubMed=9515923;
 RA Swartley J.S., Liu L.-J., Miller Y.K., Martin L.E., Edupuganti S.,
 RA Stephens D.S.;
 RT "Characterization of the gene cassette required for biosynthesis of
 the (alpha1->6)-linked N-acetyl-D-mannosamine-1-phosphate capsule of
 serogroup A *Neisseria meningitidis*.";
 RL J. Bacteriol. 180:1533-1539(1998).
 DR EMBL, AF019760; AAC38286.1; --
 SQ SEQUENCE 545 AA; 64097 MW; 1BC3400C5CC33DCD CRC64;
 Query Match 6.6%; Score 113.5; DB 2; Length 545;
 Best Local Similarity 18.4%; Pred. No. 1.1;
 Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;
 QY 3 PADSLYVY 62
 DB 395 YLNGANNAVYV 454
 QY 63 MGF-APSYVY 121
 DB 455 IANTGYLYHYH---ALISRALOS-----SDKTE 480
 QY 122 LEHMLINCSKMLPADITQLNIP--TQSSYVDPNLPVTKSLVNTCKEPTDKIKHA 180
 DB 481 LVQGNHDFK-----KLNVY 508
 QY 181 NKRFEMIEDEELAFQVIRTVNSHVVGQLDIRKNPKRVCLNDNIDHNKDAQTYKAVL 240
 ||:|||||:|:|:|


```

Db      509 -----LSVCINDGADSHLNEENVQVI- 530
QY      241 RDPFESMPFIPSOFE 255
        ||:|||||
        531 -KFELETLFPLPSSFE 544
Db
RESULT 9
Q84D00
ID      Q84D00      PRELIMINARY;      PRT;      545 AA.
AC      Q84D00;
DT      01-JUN-2003 (TREMBlrel. 24, Created)
DT      01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT      01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE      SAcB.
GN      SAcB.
OS      Neisseria meningitidis.
OC      Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OX      Neisseriaceae; Neisseria.
NCBI_TaxID=487;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=M4775;
RA      Sacchi C.T., Whitney A.A., Mayer L.W., Møthershed E., Popovic T.;
RT      "Neisseria meningitidis sacB gene.";
RL      Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY234202; AA085300.1; -.
SQ      SEQUENCE 545 AA; 64134 MW; A3648A3C85AA946F CRC64;

Query Match
Best Local Similarity 18.4%; Pred. No. 1.1;
Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

QY      3 FADSLRYVVKILNSKGFSTRKVPAPMPHMDRIYVQELQDMFPEEFDKTSFHKVHSHSD 62
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      395 YINGARANTLLEKEKFKFTTKLHTSPQSMRTDILFEMKKYPEEFNRTLHNKFRSLDD 454
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      63 MGF-AFSYFYUMLSAVQPLNISQVFDEVDTDSGVLSDREIRTLATRIHELPLSLQDLTG 121
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      455 IAVTGLYHHY-----ALLSGRALQS-----SDKTE 480
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      122 LEHMLINCSKMLPADITQANNIPP-TOESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDK 180
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      481 LVQONHDFPK-----KLNNVVTLTKERNFD-KLP----- 508
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      181 NKYRFEIMGEEIEAFKMTIRNVSHVVGQLDIRKNPKRFVCLNDINDHNKDAQTVKAVL 240
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      509 -----LSVCINDGADSHLNEENVQVI- 530
QY      241 RDPFESMPFIPSOFE 255
        ||:|||||
        531 -KFELETLFPLPSSFE 544
Db
RESULT 10
Q84C29
ID      Q84C29      PRELIMINARY;      PRT;      545 AA.
AC      Q84C29;
DT      01-JUN-2003 (TREMBlrel. 24, Created)
DT      01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT      01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE      SAcB.
GN      SAcB.
OS      Neisseria meningitidis.
OC      Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OX      Neisseriaceae; Neisseria.
NCBI_TaxID=487;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=M2677;
RA      Sacchi C.T., Whitney A.A., Mayer L.W., Møthershed E., Popovic T.;
RT      "Neisseria meningitidis sacB gene.";
RL      Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.

```

```

DR      EMBL; AY234204; AA085302.1; -.
SQ      SEQUENCE 545 AA; 64128 MW; 016B534F50D5D71 CRC64;

Query Match
Best Local Similarity 18.4%; Pred. No. 1.1;
Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

QY      3 FADSLRYVVKILNSKGFSTRKVPAPMPHMDRIYVQELQDMFPEEFDKTSFHKVHSHSD 62
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      395 YINGARANTLLEKEKFKFTTKLHTSPQSMRTDILFEMKKYPEEFNRTLHNKFRSLDD 454
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      63 MGF-AFSYFYUMLSAVQPLNISQVFDEVDTDSGVLSDREIRTLATRIHELPLSLQDLTG 121
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      455 IAVTGLYHHY-----ALLSGRALQS-----SDKTE 480
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      122 LEHMLINCSKMLPADITQANNIPP-TOESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDK 180
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      481 LVQONHDFPK-----KLNNVVTLTKERNFD-KLP----- 508
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      181 NKYRFEIMGEEIEAFKMTIRNVSHVVGQLDIRKNPKRFVCLNDINDHNKDAQTVKAVL 240
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      509 -----LSVCINDGADSHLNEENVQVI- 530
QY      241 RDPFESMPFIPSOFE 255
        ||:|||||
        531 -KFELETLFPLPSSFE 544
Db
RESULT 11
Q83U59
ID      Q83U59      PRELIMINARY;      PRT;      545 AA.
AC      Q83U59;
DT      01-JUN-2003 (TREMBlrel. 24, Created)
DT      01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      SAcB.
GN      SAcB.
OS      Neisseria meningitidis.
OC      Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OX      Neisseriaceae; Neisseria.
NCBI_TaxID=487;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=M1124, and M3562;
RA      Sacchi C.T., Whitney A.A., Mayer L.W., Møthershed E., Popovic T.;
RT      "Neisseria meningitidis sacB gene.";
RL      Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=M7060;
RA      Sacchi C.T., Whitney A.A., Mayer L.W., Popovic T.;
RT      "Neisseria meningitidis sacB gene.";
RL      Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY234203; AA085301.1; -.
DR      EMBL; AY234205; AA085303.1; -.
DR      EMBL; AY281049; AAP34772.1; -.
SQ      SEQUENCE 545 AA; 64133 MW; 7D2564286C95CD3D CRC64;

Query Match
Best Local Similarity 18.4%; Pred. No. 1.1;
Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

QY      3 FADSLRYVVKILNSKGFSTRKVPAPMPHMDRIYVQELQDMFPEEFDKTSFHKVHSHSD 62
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      395 YINGARANTLLEKEKFKFTTKLHTSPQSMRTDILFEMKKYPEEFNRTLHNKFRSLDD 454
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      63 MGF-AFSYFYUMLSAVQPLNISQVFDEVDTDSGVLSDREIRTLATRIHELPLSLQDLTG 121
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      455 IAVTGLYHHY-----ALLSGRALQS-----SDKTE 480
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      122 LEHMLINCSKMLPADITQANNIPP-TOESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDK 180
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      481 LVQONHDFPK-----KLNNVVTLTKERNFD-KLP----- 508
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

QY 181 NKYRPEIMGEELAFKMTITNVSHVVGQLDIRKPKRVCUNIDHNHDKAQYKAVL 240
DB 509 -----LSCVINCDAHSHLNEEMVQVI- 530

QY 241 RDPYSEMPPIPSQFB 255
DB 531 -KFLETLFPLPSSFE 544

RESULT 12
QJQJWB
ID QJQJWB PRELIMINARY; PRT; 545 AA.
AC QJQJWB;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Putative capsule biosynthesis protein.
GN SABC OR NMA0200.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=62491 / Serogroup A / Serotype 4A;
RA MEDLINE=2022256; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Bellwell T., Hamlin N., Holroyd S.,
RA Tagatz K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajadurai M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506 (2000).
DR EMBL; AL162752; CAB83514.1;
DR PIR; C82014; C82014.
KW Complete proteome.
SQ SEQUENCE 545 AA; 64133 MW; 7D2564286C95CD3D CRC64;

Query Match
Best Local Similarity 18.4%; Pred. No. 1.1;
Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

QY 3 FADSLRYVKNLINSKFGFTSKVPAHMPHMDIRYMOELQMFPEEPKTSFHKVRSSE 62
DB 395 YLNGARNANTLLEKEFKFTTKLHTSPQSWRTDILFEMEKCYPEEFNRRLHNKFRSLDD 454
QY 63 MGF-APSYEYLMASVQPLNISQVDFDEVTDQSGVLSPREIRTLATRIHELPLSLQDLTG 121
DB 455 IAVTGLYHNY-----ALLSGRALQS-----SDKTE 480
QY 122 LEHMLINCSKMLPADITQNNIP-TPQSSYDPLNIPYTKSLVTNCKEPTDKIRKAYKDK 180
DB 481 LVQCHNDPFK-----KLNNVTLTKERNFD-KLP----- 508
QY 181 NKYRPEIMGEELAFKMTITNVSHVVGQLDIRKPKRVCUNIDHNHDKAQYKAVL 240
DB 509 -----LSCVINCDAHSHLNEEMVQVI- 530

QY 241 RDPYSEMPPIPSQFB 255
DB 531 -KFLETLFPLPSSFE 544

RESULT 13
Q26023 PRELIMINARY; PRT; 1048 AA.
AC Q26023;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

```

```

DE Hypothetical protein (Fragment).
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Alano P., Elliott J.F.;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; M69147; AAA74655.1; -.
KW Hypothetical protein.
FT NON TER 1
FT NON TER 1048
SQ SEQUENCE 1048 AA; 126518 MW; 7AF051480FA22424 CRC64;

Query Match
Best Local Similarity 22.7%; Pred. No. 3.9;
Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;

QY 41 LQMFPEEDKTSFHKVRSSEDMQFAFSYFYILMSAVQPLNISQVDFDEVTDQSGVLS 99
DB 592 LEDIYVE---KKKYKBEYSKMRRISSMLDYEVNK---QIKENYHKVDT-----ISBH 639
QY 100 --REIR-----TLATIHLEPLSLQ-----DITGLEHMLINCSKMLPADITQNNIP-- 145
DB 640 KFOEIRQHRDKLIENTHELKYMVYQIOLITNYHOLENHSLELQALQOKNIPRL 699
QY 146 -----TQSSYDPLNIPV-TKSLVTCRPTVDKIRKAYKDK-----KMKYRPEI-- 187
DB 700 NVLEKLEITTKR--KKNPDIISTSHARDQGVSDTLRGANHGDIKGDNDNEVLLI 757
QY 188 -----MGEELIAFKIRITNVSHVVCQ-----LDDIRKPKRVCUNIDHN 228
DB 758 EQIQSLTKRWGNQGVGSLERKNNLSQYQYLODKLNVVDIYKNNF--KHYLEK 814
QY 229 NHHDAO-----TVAVNRDPYSEMPPIPSQFELPREYRNFLEHGELOEYRAYRDKLK 281
DB 815 LKESKINREKFTTKVDVLSNYSTL-----EYMWKFL-LHSDQESFEEKLE 862

RESULT 14
Q25857 PRELIMINARY; PRT; 3119 AA.
ID Q25857;
AC Q25857;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Pfg377.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA MEDLINE=9630472; PubMed=8719156;
RA Alano P., Read D., Bruce M., Aikawa M., Kaide T., Tegoshi T.,
RA Bhatti S., Smith D.K., Luo C., Hanera S., Carter R., Elliott J.F.;
RT "COS cell expression cloning of Pfg377, a Plasmodium falciparum
RT gametocyte antigen associated with oocystophilic bodies."
RL KOL. Biochem. Parasitol. 74:143-156 (1995).
DR EMBL; L04161; AAC37257.1; -.
DR PIR; T18414; T18414.
DR InterPro; IPR001064; Crystal11n.
DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA. 1.
SQ SEQUENCE 3119 AA; 377353 MW; 519B9D5BDBFCRC CRC64;

Query Match
Best Local Similarity 6.4%; Score 111; DB 5; Length 3119;
Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;

QY 41 LQMFPEEDKTSFHKVRSSEDMQFAFSYFYILMSAVQPLNISQVDFDEVTDQSGVLS 99
DB 1261 LEDIYVE---KKKYKBEYSKMRRISSMLDYEVNK---QIKENYHKVDT-----ISBH 1308

```

QY 100 --REIR-----TLATRIHELPLSLQ-----DITGLEHMLINCSKMLPADITQNLNIP-- 145
 Db 1309 KFOELRQHRDKIENTHELYKEMVVOIQIDLTNYHOLENISELLOALQONKNIPRHL 1368
 QY 146 -----TQESYDPLNPV-TKSLVTNCKPVTDKIHAKYD-----KMKRPEI--- 187
 Db 1369 NVLEKKLEITTKR--KKKKPDISTSSHATDEQVSDTLIRGAHNGDIIKGEDNDEVLLI 1426
 QY 188 -----MGEEBIAFKIRTNVSHVVGQ-----LDDIRKPPKFCVCLNDNIDH 228
 Db 1427 EQIQSLKTMGDNQVGSILEKLNLSPOYQLODKXNVVEDIYKNLRNF--KHVYEK 1483
 QY 229 NHKDAO-----TYKAVLRDPFYESMFPISQFELPREYRNFELHMEHLOEWRAYRDKLK 281
 Db 1484 LHKESKINREKITKYDVLSNVYSTL-----EYWKFL-LHDFQWSEFKDELE 1531

RESULT 15

0814T3
 ID 0814T3 PRELIMINARY; PRT; 3119 AA.
 AC 0814T3
 DT 01-MAR-2003 (TEMBLrel. 23, Created)
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE PFG377 protein.
 GN PFL2405C.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_Taxid=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=2255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Bertman M., Hyman R.W.,
 RA Carlton J.M., Patil A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
 RA Berthe M., Allen V., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McFadden G.I., Cummings L.W., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrall B.,
 RA "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum.";
 RL Nature 419:498-511(2002).
 DR EMBL; AE014852; AAN36565.1; "-
 DR InterPro; IPR001064; Crystal1in.
 DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.
 SQ SEQUENCE 3119 AA; 377380 MW; 3A6B67F1CD5C1CF CRC64;

Query Match 6.4%; Score 111; DB 5; Length 3119;
 Best Local Similarity 22.7%; Pred. No. 15;
 Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;

QY 41 LODMFEEDKTSFHKHSEDMQFAFSFYTLMSAVQPLNISQVDFDEVDTQDSGLSD- 99
 Db 1261 LEDITWE--KKKYYKEYSKKRRIISSNLDYVNR--QIKHYHKVDT-----ISEH 1308
 QY 100 --REIR-----TLATRIHELPLSLQ-----DITGLEHMLINCSKMLPADITQNLNIP-- 145
 Db 1309 KFOELRQHRDKIENTHELYKEMVVOIQIDLTNYHOLENISELLOALQONKNIPRHL 1368
 QY 146 -----TQESYDPLNPV-TKSLVTNCKPVTDKIHAKYD-----KMKRPEI--- 187
 Db 1369 NVLEKKLEITTKR--KKKKPDISTSSHATDEQVSDTLIRGAHNGDIIKGEDNDEVLLI 1426
 QY 188 -----MGEEBIAFKIRTNVSHVVGQ-----LDDIRKPPKFCVCLNDNIDH 228
 Db 1427 EQIQSLKTMGDNQVGSILEKLNLSPOYQLODKXNVVEDIYKNLRNF--KHVYEK 1483
 QY 229 NHKDAO-----TYKAVLRDPFYESMFPISQFELPREYRNFELHMEHLOEWRAYRDKLK 281

Db 1484 LHKESKINREKITKYDVLSNVYSTL-----EYWKFL-LHDFQWSEFKDELE 1531
 Search completed: July 26, 2004, 11:11:37
 Job time : 23.4324 secs

This Page Blank (uspto)

CC alpha-glucosidase (N-acetylglucosamine-1-phosphodiester alpha-N-
CC Acetylglucosaminidase) (II). The protein of the invention has
CC nephrotoxic activity, and may be useful in enzyme replacement therapy. A
CC protein of the invention (I), (II) is useful for preparing a
CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase
CC comprising a terminal mannose-6-phosphate, is useful for treating a
CC patient suffering from a lysosomal storage disease. The present sequence
CC is used in the exemplification of the invention

XX Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLILGLSGAPAPAGAAKMKVVEBNPFGVNNPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLILGLSGAPAPAGAAKMKVVEBNPFGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVLFLSLGKCSLVSTYKYFCPPHNTQHEOTFRMNAVSGIIGIHWHEIANNPTG 120
DB 61 PVLFLSLGKCSLVSTYKYFCPPHNTQHEOTFRMNAVSGIIGIHWHEIANNPTG 120
QY 121 MMRDGDACRSRSGKSVLACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVPTL 180
DB 121 MMRDGDACRSRSGKSVLACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVPTL 180
QY 181 PEARLOROMQVQEDLADDELITPQGHKRLRTLFEAGYLKTPBENBPQLGSPDLSGFE 240
DB 181 PEARLOROMQVQEDLADDELITPQGHKRLRTLFEAGYLKTPBENBPQLGSPDLSGFE 240
QY 241 TLNCRKAKHKEISKEIKRLKGLLTQHGIPYTPETSNLEHLGHETPRKSPBQLRGDPG 300
DB 241 TLNCRKAKHKEISKEIKRLKGLLTQHGIPYTPETSNLEHLGHETPRKSPBQLRGDPG 300
QY 301 LKGS 305
DB 301 LKGS 305

RESULT 2
ABG92075
ID ABG92075 standard; protein; 305 AA.

XX AC ABG92075;
XX DT 29-NOV-2002 (first entry)

DE Human receptors and membrane associated protein REMAP-18.

XX Human; receptor and membrane-associated protein; REMAP; atherosclerosis;
KW cardiovascular disorder; hypertension; congestive heart failure; edema;
KW aneurysm; angina pectoris; ischaemic heart disease; lung disease; nausea;
KW rheumatic heart disease; chronic obstructive pulmonary disease; anaemia;
KW emphysema; bronchitis; gastrointestinal disorder; peptic ulcer; asthma;
KW Crohn's disease; lipid metabolism disorder; Fabry's disease; dementia;
KW pancreatitis; diabetes mellitus; hyperlipidaemia; autoimmune disorder;
KW inflammatory disorder; neurological disorder; Creutzfeldt-Jakob disease;
KW cell proliferative disorder; protein replacement therapy; adenocarcinoma;
KW developmental disorder; metabolic disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; multiple sclerosis; Cushing's syndrome; anxiety;
KW schizophrenia; Addison's disease; endocrine disorder; gene therapy; gout;
KW amnesia; cancer; leukaemia; lymphoma; myeloma; sarcoma.

OS Homo sapiens.

XX WO200263006-A2.

XX 15-AUG-2002.

XX 05-FEB-2002; 2002WO-US003866.

XX 06-FEB-2001; 2001US-0267201P.

PR 16-FEB-2001; 2001US-0269580P.
PR 03-APR-2001; 2001US-0282679P.
PR 02-MAY-2001; 2001US-0288295P.
PR 14-JAN-2002; 2002US-0348687P.

PA (INCY-) INCYTE GENOMICS INC.

XX Airmatzi Y, Yue H, Ding L, Nguyen DB, Gandhi AR, Burford N;
PI Thangavelu K, Elliott VS, Ramkumar U, Yao MG, Lal PG, Tang TY,
PI Swarnakar A, Warren BA, Walla NK, Policky JL, Xu Y, Horneill CD,
PI Au-Yang J, Baughn WR, Duggan BM, Lu DAM, Gietzen KU, Hillmann JL;
PI Raumann BE, Lu Y, Sappenstein SK, Tran UK, Richardson TW,
PI Emerling BM, Hafia A, Burrill JD, Marcus GA, Zingler KA,
PI Kable AE, Gorrard AE;

XX WPI, 2002-627559/67.
DR N-PSDB; ABS67796.

PT New human receptor and membrane-associated proteins and polynucleotides,
PT useful for diagnosing, treating or preventing cardiovascular disorders
PT (e.g. aneurysms), neurological disorders (e.g. Parkinson's disease) or
PT cancers.

XX Claim 2; Page 185-186; 262pp; English.

XX The present invention relates to a new receptor and membrane-associated
CC protein (REMAP). The polypeptide, polynucleotide and agonist are useful
CC for treating a condition associated with decreased expression of
CC functional REMAP. The antagonist is useful for treating a disease
CC associated with overexpression of functional REMAP. The anti-REMAP
CC antibody is useful for diagnosing a condition or disease associated with
CC the expression of REMAP. These polypeptides, polynucleotides, agonists
CC and antagonists are particularly useful for diagnosing, treating or
CC preventing cardiovascular (e.g. atherosclerosis, hypertension, aneurysms,
CC congestive heart failure, angina pectoris, or ischemic or rheumatic
CC heart disease), lung (e.g. edema, chronic obstructive pulmonary disease,
CC emphysema or bronchitis) gastrointestinal (e.g. nausea, peptic ulcer or
CC Crohn's disease), lipid metabolism (e.g. Fabry's disease, diabetes,
CC mellitus or hyperlipidaemia), autoimmune/inflammatory (e.g. anaemia,
CC asthma, gout, pancreatitis or Crohn's disease), neurological (e.g.
CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
CC multiple sclerosis, Parkinson's disease, anxiety, schizophrenia or
CC amnesia), metabolic (e.g. Addison's disease), developmental (e.g.
CC Cushing's syndrome), endocrine or cell proliferative disorders (e.g.
CC cancers including adenocarcinoma, leukaemia, lymphoma, myeloma or
CC sarcoma). The present amino acid sequence represents a human REMAP
CC protein of the invention

XX Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 5; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLILGLSGAPAPAGAAKMKVVEBNPFGVNNPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLILGLSGAPAPAGAAKMKVVEBNPFGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVLFLSLGKCSLVSTYKYFCPPHNTQHEOTFRMNAVSGIIGIHWHEIANNPTG 120
DB 61 PVLFLSLGKCSLVSTYKYFCPPHNTQHEOTFRMNAVSGIIGIHWHEIANNPTG 120
QY 121 MMRDGDACRSRSGKSVLACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVPTL 180
DB 121 MMRDGDACRSRSGKSVLACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVPTL 180
QY 181 PEARLOROMQVQEDLADDELITPQGHKRLRTLFEAGYLKTPBENBPQLGSPDLSGFE 240
DB 181 PEARLOROMQVQEDLADDELITPQGHKRLRTLFEAGYLKTPBENBPQLGSPDLSGFE 240
QY 241 TLNCRKAKHKEISKEIKRLKGLLTQHGIPYTPETSNLEHLGHETPRKSPBQLRGDPG 300
DB 241 TLNCRKAKHKEISKEIKRLKGLLTQHGIPYTPETSNLEHLGHETPRKSPBQLRGDPG 300

QY 301 LRGS 305
 DB 301 LRGS 305

RESULT 3
 ADA57553
 ID ADA57553 standard; protein; 305 AA.
 AC ADA57553;
 DT 20-NOV-2003 (first entry)
 DE Human secreted protein #500.

XX immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;
 KM cytototoxic; cerebroprotective; neuroprotective; nootropic;
 KM cardiovascular; antiarteriosclerotic; gene therapy;
 KM human secreted protein; immune disorder; inflammation;
 KM inflammatory bowel disease; cancer; CNS disorder; neurodegenerative disorders;
 KM multiple sclerosis; ischaemic brain injury; Parkinson's disease; asthma; allergy;
 KM Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
 KM triple helix formation; antisense gene therapy; forensic biology.

OS Homo sapiens.
 XX
 PN WO2002102994-A2.
 PD 27-DEC-2002.
 XX
 PF 19-MAR-2002; 2002WO-US008278.
 XX
 PR 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX WPI; 2003-167512/16.
 DR N-PSDB; ADA56660.

XX New human secreted polypeptides and polynucleotides, useful for
 PT diagnosing, treating or preventing e.g. immune disorders, inflammatory
 PT conditions, respiratory disorders, cancers, CNS disorders, or
 PT neurodegenerative disorders.

XX Claim 13; SEQ ID NO 1746; 1754pp; English.

XX The invention relates to 592 new human secreted polypeptides useful for
 CC diagnosing, treating or preventing e.g. immune disorders, inflammatory
 CC conditions, respiratory disorders, cancers, CNS disorders, or
 CC neurodegenerative disorders, or polypeptides comprising an amino acid
 CC sequence at least 95% identical to the new sequences. The polypeptides,
 CC antibodies or antibody fragments that bind to the polypeptides, nucleic
 CC acids encoding the polypeptides, agonists or antagonists that binds to
 CC the polypeptide, are useful in preparing diagnostic or pharmaceutical
 CC compositions for diagnosing, treating or preventing an e.g. immune
 CC disorders, inflammatory conditions (e.g. inflammatory bowel disease,
 CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
 CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
 CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
 CC disorders (e.g. Parkinson's disease or Alzheimer's disease), and
 CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
 CC polynucleotides are useful for chromosome identification, chromosome
 CC mapping, for controlling gene expression through triple helix formation
 CC or antisense DNA or RNA, in gene therapy, for identifying individuals
 CC from minute biological samples, in forensic biology, and as hybridization
 CC probes. The polypeptides are useful for as molecular weight markers on
 CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

CC gels, to raise antibodies, for testing biological activities, and for
 CC treating or preventing neural disorders, immune system disorders,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, proliferative and/or cancerous diseases. This sequence corresponds
 CC to one of the polypeptide of the invention. Note: The sequence data for
 CC this patent did form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 305 AA;
 SQ

Query Match 100.0%; Score 1634; DB 6; Length 305;
 Best Local Similarity 100.0%; Pred. No. 3.2e-16;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLGLSAGPAPGAAAKKVEEPNAPGVNNPPLPQASRLQAKRDPSPVSG 60
 DB 1 MAAGLARLLLLGLSAGPAPGAAAKKVEEPNAPGVNNPPLPQASRLQAKRDPSPVSG 60

QY 61 PVHLFPLSGKCSLYSTKYECFPPHVTQHEQTFRNVAAGIIGIHWEMETIANTFTG 120
 DB 61 PVHLFPLSGKCSLYSTKYECFPPHVTQHEQTFRNVAAGIIGIHWEMETIANTFTG 120

QY 121 MMWRDGDACRSRSRQSGVZLACGKSNRLAHSSEPTCYVALFETPLVCHPHALLVYPTL 180
 DB 121 MMWRDGDACRSRSRQSGVZLACGKSNRLAHSSEPTCYVALFETPLVCHPHALLVYPTL 180

QY 181 PEALQROMQVQEDDLADELITPOGHEKILRTLPEQGYKTEBENPPQLBGGPSLIGE 240
 DB 181 PEALQROMQVQEDDLADELITPOGHEKILRTLPEQGYKTEBENPPQLBGGPSLIGE 240

QY 241 TLENCRAKHELSKEIKRLKGLLTQHGIPYTPTEFSNLEHLGHETPRAKSPBOLRGDPG 300
 DB 241 TLENCRAKHELSKEIKRLKGLLTQHGIPYTPTEFSNLEHLGHETPRAKSPBOLRGDPG 300

QY 301 LRGS 305
 DB 301 LRGS 305

RESULT 4
 ADA57217
 ID ADA57217 standard; protein; 305 AA.
 AC ADA57217;
 DT 20-NOV-2003 (first entry)
 DE Human secreted protein #500.

XX immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;
 KM cytototoxic; cerebroprotective; neuroprotective; nootropic;
 KM cardiovascular; antiarteriosclerotic; gene therapy;
 KM human secreted protein; immune disorder; inflammation;
 KM respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;
 KM inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;
 KM multiple sclerosis; ischaemic brain injury; Parkinson's disease;
 KM Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
 KM triple helix formation; antisense gene therapy; forensic biology.

OS Homo sapiens.
 XX
 PN WO2002102994-A2.
 PD 27-DEC-2002.
 XX
 PF 19-MAR-2002; 2002WO-US008278.
 XX
 PR 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;
XX MPI: 2003-167512/16.
DR N-PSDB; ADA56321.
XX
XX New human secreted polypeptides and polynucleotides, useful for
PT diagnosing, treating or preventing e.g. immune disorders, inflammatory
PT conditions, respiratory disorders, cancers, CNS disorders, or
PT neurodegenerative disorders.
XX
PS Claim 13; SEQ ID NO 1407; 1754bp; English.
XX
XX The invention relates to 592 new human secreted polypeptides useful for
CC diagnosing, treating or preventing e.g. immune disorders, inflammatory
CC conditions, respiratory disorders, cancers, CNS disorders, or
CC neurodegenerative disorders, or polypeptides comprising an amino acid
CC sequence at least 95% identical to the new sequences. The polypeptides,
CC antibodies or antibody fragments that bind to the polypeptides, nucleic
CC acids encoding the polypeptides, agonists or antagonists that binds to
CC the polypeptide, are useful in preparing diagnostic or pharmaceutical
CC compositions for diagnosing, treating or preventing an e.g. immune
CC disorders, inflammatory conditions (e.g. inflammatory bowel disease,
CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
CC disorders (e.g. Parkinson's disease or Alzheimer's disease), and
CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
CC polynucleotides are useful for chromosome identification, chromosome
CC mapping, for controlling gene expression through triple helix formation
CC or antisense DNA or RNA, in gene therapy, for identifying individuals
CC from minute biological samples, in forensic biology, and as hybridization
CC probes. The polypeptides are useful for as molecular weight markers on
CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
CC gels, to raise antibodies, for testing biological activities, and for
CC treating or preventing neural disorders, immune system disorders,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, proliferative and/or cancerous diseases. This sequence corresponds
CC to one of the polypeptide of the invention. Note: The sequence data for
CC this patent did form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 305 AA;
Query Match 100.0%; Score 1634; DB 6; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NAAGLRLLLILGLSAGGAPAGAAKMKVEEPNFAFGVNNPFLPOASRLQAKRDPSPVSG 60
DB 1 NAAGLARLLLLGLSAGGAPAGAAKMKVEEPNFAFGVNNPFLPOASRLQAKRDPSPVSG 60
QY 61 PHLRLSLGSKCSLVESYTKYKFCFPHNVTOHEQTFRRNNAVSGILGHEWEIANNFTFG 120
DB 61 PHLRLSLGSKCSLVESYTKYKFCFPHNVTOHEQTFRRNNAVSGILGHEWEIANNFTFG 120
QY 121 MMNRGDACRSRSPSKVELACGKSNRLAHVSEPTCYATLTPETPLVCHHALVPTL 180
DB 121 MMNRGDACRSRSPSKVELACGKSNRLAHVSEPTCYATLTPETPLVCHHALVPTL 180
QY 181 PEALORQWQVQDIADDELITPQGEKLLRTLFEEDAGYLKTPPENEPTOLEGGPDSIGFE 240
DB 181 PEALORQWQVQDIADDELITPQGEKLLRTLFEEDAGYLKTPPENEPTOLEGGPDSIGFE 240
QY 241 TLNCRKXAKHLSKSLKRLKGLLTQHGIPYTRPTSTLHLGHETPRAPKSPBOLRGDPG 300
DB 241 TLNCRKXAKHLSKSLKRLKGLLTQHGIPYTRPTSTLHLGHETPRAPKSPBOLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305

RESULT 5
ID ADA41096
XX ADA41096 standard; protein; 305 AA.
XX
XX ADA41096;
AC
XX 20-NOV-2003 (first entry)
DT
XX
DE Human secreted protein.
XX
XX Human, secreted protein; cancer; hyperproliferative disorder;
KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
KW anaemia; allergic reaction; asthma; cardiovascular disorder;
KW wound healing; cytotoxic; immunosuppressive; neotropic; neuroprotective;
KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
KW vulneary; caridiant; gene therapy.
XX
XX Homo sapiens.
OS
XX
XX WO2002102993-A2.
PN
XX
XX 27-DEC-2002.
PD
XX
XX 19-MAR-2002; 2002WO-US008123.
PF
XX
XX 21-MAR-2001; 2001US-0277340P.
PR
XX 19-JUL-2001; 2001US-0306171P.
PR
XX 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
PI
XX
XX MPI: 2003-175238/17.
DR
XX
XX New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating cancer or other hyperproliferative disorder,
PT asthma, allergies or AIDS.
XX
XX
PS Claim 1; SEQ ID NO 1478; 3205bp; English.
XX
XX The invention relates to novel genes ADA39629-ADA40565 and proteins
CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
CC fragments, and agonists or antagonists that bind to the polypeptide are
CC useful for preparing a diagnostic or pharmaceutical composition for
CC diagnosing or treating cancer or other hyperproliferative disorder. The
CC polypeptides and nucleic acid molecules are also useful for detecting,
CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC or other hyperproliferative disorders including neoplasms, autoimmune
CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC thrombocytopenia), allergic reactions including asthma or eczema,
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC fungal or viral infections including HIV/AIDS), or wound healing and
CC disorders of epithelial cell proliferation. The nucleic acids are also
CC useful for chromosome identification, radiation hybrid mapping or long-
CC range restriction mapping, as molecular weight markers, or as
CC hybridization or diagnostic probes. The polypeptides and antibodies are
CC useful for providing immunological probes for differential identification
CC of the tissues immunohistochemistry assays. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 6; Length 305;
 Best Local Similarity 100.0%; Pred. No. 3.2e-161;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARILLILGLSAGGPAPAGAAKMKVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
 DB 1 MAAGLARILLILGLSAGGPAPAGAAKMKVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
 QY 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNVTOHEQTFRNNAVSGILGIWHEMEIANNTFTG 120
 DB 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNVTOHEQTFRNNAVSGILGIWHEMEIANNTFTG 120
 QY 121 MMNRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFEPPLVCHPALLVYPTL 180
 DB 121 MMNRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFEPPLVCHPALLVYPTL 180
 QY 181 PEALQROMDQVEODLDELITPOGHEKILRTLFEAGYLTKEPENEPTQLBGGPDSIGFE 240
 DB 181 PEALQROMDQVEODLDELITPOGHEKILRTLFEAGYLTKEPENEPTQLBGGPDSIGFE 240
 QY 241 TLNCRKAHKELSKEIKRLKGLLTGIGIPYTRPETSNTLEHGHETPRAKSPBQLRGDPG 300
 DB 241 TLNCRKAHKELSKEIKRLKGLLTGIGIPYTRPETSNTLEHGHETPRAKSPBQLRGDPG 300
 QY 301 LRGS 305
 DB 301 LRGS 305

RESULT 6
 ADA41445
 ID ADA41445 standard; protein; 305 AA.
 AC ADA41445;
 DT 20-NOV-2003 (first entry)
 DE Human secreted protein.
 DE Human; secreted protein; cancer; hyperproliferative disorder;
 KM rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
 KM anaemia; allergic reaction; asthma; cardiovascular disorder;
 KM wound healing; cytostatic; immunosuppressive; nocotropic; neuroprotective;
 KM antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
 KM vulnerability; cardiac; gene therapy.
 OS Homo sapiens.
 XX WO2002102993-A2.
 XX 27-DEC-2002.
 PD 19-MAR-2002; 2002WO-US008123.
 PF 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM;
 PI WPI; 2003-175238/17.
 DR New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,
 PT preventing or treating cancer or other hyperproliferative disorder,
 PT asthma, allergies or AIDS.
 XX Claim 1; SEQ ID NO 1828; 3205bp; English.
 PS The invention relates to novel genes ADA39629-ADA40565 and proteins

CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
 CC fragments, and agonists or antagonists that bind to the polypeptide are
 CC useful for preparing a diagnostic or pharmaceutical composition for
 CC diagnosing or treating cancer or other hyperproliferative disorder. The
 CC polypeptides and nucleic acid molecules are also useful for detecting,
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
 CC or other hyperproliferative disorders including neoplasms, autoimmune
 CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
 CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
 CC thrombocytopenia), allergic reactions including asthma or eczema,
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
 CC fungal or viral infections including HIV/AIDS), or wound healing and
 CC disorders of epithelial cell proliferation. The nucleic acids are also
 CC useful for chromosome identification, radiation hybrid mapping or long-
 CC range restriction mapping, as molecular weight markers, or as
 CC hybridization or diagnostic probes. The polypeptides and antibodies are
 CC useful for providing immunological probes for differential identification
 CC of the tissues immunohistochemistry assays. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 305 AA;
 QY 1 MAAGLARILLILGLSAGGPAPAGAAKMKVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
 DB 1 MAAGLARILLILGLSAGGPAPAGAAKMKVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
 QY 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNVTOHEQTFRNNAVSGILGIWHEMEIANNTFTG 120
 DB 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNVTOHEQTFRNNAVSGILGIWHEMEIANNTFTG 120
 QY 121 MMNRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFEPPLVCHPALLVYPTL 180
 DB 121 MMNRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFEPPLVCHPALLVYPTL 180
 QY 181 PEALQROMDQVEODLDELITPOGHEKILRTLFEAGYLTKEPENEPTQLBGGPDSIGFE 240
 DB 181 PEALQROMDQVEODLDELITPOGHEKILRTLFEAGYLTKEPENEPTQLBGGPDSIGFE 240
 QY 241 TLNCRKAHKELSKEIKRLKGLLTGIGIPYTRPETSNTLEHGHETPRAKSPBQLRGDPG 300
 DB 241 TLNCRKAHKELSKEIKRLKGLLTGIGIPYTRPETSNTLEHGHETPRAKSPBQLRGDPG 300
 QY 301 LRGS 305
 DB 301 LRGS 305

RESULT 7
 ADB91814
 ID ADB91814 standard; protein; 305 AA.
 AC ADB91814;
 DT 04-DEC-2003 (first entry)
 DE Human secreted protein #SEQ ID 760.
 DE Human secreted protein; gene therapy; antidiabetic; diabetes; human.
 OS Homo sapiens.
 XX

PN WO2003004622-A2.
XX
PD 16-JAN-2003.
XX
PF 19-MAR-2002; 2002WO-US008124.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-229407/22.
XX
PT Nucleic acid encoding a human secreted protein is useful in diagnosing or
XX treating diabetes or conditions related to diabetes.
XX
PS Claim 3; SEQ ID NO 760; 1537bp; English.
XX
CC The invention relates to isolated nucleic acid molecules ADB91065-
CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
CC ADB91834. Also disclosed is a recombinant vector comprising a
CC polynucleotide of the invention, and a recombinant host cell comprising
CC the recombinant vector. The polypeptide of the invention is useful in
CC identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases activity of the polypeptide. The polypeptide, polynucleotide,
CC antibody or its fragment, agonist or antagonist are useful for preparing
CC a pharmaceutical composition for diagnosing or treating diabetes or
CC conditions related to diabetes. The present sequence is that of the human
CC immunoglobulin Fc portion used to generate fusion proteins, increasing
CC the stability of the fused protein as compared to the secreted protein
CC only. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 305 AA;
XX
Query Match 100.0%; Score 1634; DB 7; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEBPNAFGVNNPPLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEBPNAFGVNNPPLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRNMAYSGLIGIHEWEIANTPTG 120
DB 61 PVHLFRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRNMAYSGLIGIHEWEIANTPTG 120
QY 121 MMRRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVPTL 180
DB 121 MMRRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVPTL 180
QY 181 PEALOROMQOVQODLADBLITPQGHKILRTLFEDAGYUKTPEENEPQLQEGSPSLGFE 240
DB 181 PEALOROMQOVQODLADBLITPQGHKILRTLFEDAGYUKTPEENEPQLQEGSPSLGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPPAKSPQOLRGDPG 300
DB 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPPAKSPQOLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305
RESULT 8
ADB91665
ID ADB91665 standard; protein; 305 AA.
XX

AC ADB91665;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human secreted protein #SEQ ID 611.
XX
KM Secreted protein; gene therapy; antidiabetic; diabetes; human.
XX
OS Homo sapiens.
XX
PN WO2003004622-A2.
XX
PD 16-JAN-2003.
XX
PF 19-MAR-2002; 2002WO-US008124.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-229407/22.
XX
PT Nucleic acid encoding a human secreted protein is useful in diagnosing or
XX treating diabetes or conditions related to diabetes.
XX
PS Claim 3; SEQ ID NO 611; 1537bp; English.
XX
CC The invention relates to isolated nucleic acid molecules ADB91065-
CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
CC ADB91834. Also disclosed is a recombinant vector comprising a
CC polynucleotide of the invention, and a recombinant host cell comprising
CC the recombinant vector. The polypeptide of the invention is useful in
CC identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases activity of the polypeptide. The polypeptide, polynucleotide,
CC antibody or its fragment, agonist or antagonist are useful for preparing
CC a pharmaceutical composition for diagnosing or treating diabetes or
CC conditions related to diabetes. The present sequence is that of the human
CC immunoglobulin Fc portion used to generate fusion proteins, increasing
CC the stability of the fused protein as compared to the secreted protein
CC only. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 305 AA;
XX
Query Match 100.0%; Score 1634; DB 7; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEBPNAFGVNNPPLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEBPNAFGVNNPPLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRNMAYSGLIGIHEWEIANTPTG 120
DB 61 PVHLFRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRNMAYSGLIGIHEWEIANTPTG 120
QY 121 MMRRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVPTL 180
DB 121 MMRRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVPTL 180
QY 181 PEALOROMQOVQODLADBLITPQGHKILRTLFEDAGYUKTPEENEPQLQEGSPSLGFE 240
DB 181 PEALOROMQOVQODLADBLITPQGHKILRTLFEDAGYUKTPEENEPQLQEGSPSLGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPPAKSPQOLRGDPG 300
DB 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPPAKSPQOLRGDPG 300

QY 301 LRGS 305
Db 301 LRGS 305

RESULT 9
ADCT4325
ID ADCT4325 strand: protein; 305 AA.
AC ADCT4325;
DT 01-JAN-2004 (first entry)
DE Human secreted protein - SEQ ID 958.

antianemic; antirheumatic; antiarthritic; antiinflammatory; antithyroid;
antidiabetic; immunosuppressive; dermatologic; nephrotoxic;
antiparkinsonian; neuroprotective; nootropic; antibacterial; virucide;
fungicide; antiparasitic; antiarteriosclerotic; vulnery; cytostatic;
haemopoietic; haematologic; anaemia; autoimmune disorder;
rheumatoid arthritis; inflammation; Grave's disease; diabetes;
systemic lupus erythematosus; glomerulonephritis; neurodegenerative;
Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;
cancer; bacterial; viral; fungal; parasitic infection; gene therapy;
human.

OS Homo sapiens.
XX WO2003038063-A2.
XX 08-MAY-2003.
PD 19-MAR-2002; 2002WO-US008277.
XX 21-MAR-2001; 2001US-0277340P.
XX 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX MPI: 2003-430516/40.
XX N-PSDB; ADCT3710.
DR

New human secreted polypeptide for diagnosing, preventing or treating
PT hematopoietic or hematologic disorders (e.g. anemia), autoimmune
PT disorders (e.g. diabetes) or hyperproliferative disorders (e.g. cancer or
PT atherosclerosis).

PS Claim 16; SEQ ID NO 958; 2272BP; English.

The invention relates to a novel human secreted polypeptide comprising a
CC defined sequence given in the specification. The polypeptide, nucleic
CC acid molecule, antibody, agonist or antagonist of the invention may be
CC useful for preparing a composition for diagnosing or treating a
CC haemopoietic or haematologic disorder such as anaemia, autoimmune
CC disorders such as rheumatoid arthritis, inflammation, Grave's disease,
CC diabetes, systemic lupus erythematosus or glomerulonephritis,
CC neurodegenerative disorders including Parkinson's disease and Alzheimer's
CC disease, wounds and hyperproliferative disorders including
CC atherosclerosis or cancer, as well as bacterial, viral, fungal or
CC parasitic infections. The polypeptide may also be used during gene
CC therapy procedures and for identifying a binding partner by contacting
CC the polypeptide with a binding partner and determining whether the
CC binding partner increases or decreases the activity of the polypeptide.
CC The current sequence is that of the human secreted protein of the
CC invention.

XX Sequence 305 AA;
SQ

Query Match 100.0%; Score 1634; DB 7; Length 305;

Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARIILLILGLSAGGAPAGAAKXKVEEENAFGVNPPFLPQASRLQAKRDPSPVSG 60
Db 1 MAAGLARIILLILGLSAGGAPAGAAKXKVEEENAFGVNPPFLPQASRLQAKRDPSPVSG 60

QY 61 PVHLFRLSGKCFSTLVSTKYERCFPHNVOHQOTFRMAVSGILGIMHEMELANNTPFG 120
Db 61 PVHLFRLSGKCFSTLVSTKYERCFPHNVOHQOTFRMAVSGILGIMHEMELANNTPFG 120

QY 121 MMRDGDACRSRSGQKVELACGKSNRLAHVSEPTCYVALTFETPELVCHPALLVPTL 180
Db 121 MMRDGDACRSRSGQKVELACGKSNRLAHVSEPTCYVALTFETPELVCHPALLVPTL 180

QY 181 PEALQOMQOVEDDLADELITPGHEKRLRTLPEDGVLKTPBEENPTQLGGPDLGFE 240
Db 181 PEALQOMQOVEDDLADELITPGHEKRLRTLPEDGVLKTPBEENPTQLGGPDLGFE 240

QY 241 TLENCKRAHKEJSKEIKRLKGLLTQHGIPYTRPTERSNLEHGHETPRAKSPBQLRGDPG 300
Db 241 TLENCKRAHKEJSKEIKRLKGLLTQHGIPYTRPTERSNLEHGHETPRAKSPBQLRGDPG 300

QY 301 LRGS 305
Db 301 LRGS 305

RESULT 10
ADCT4565
ID ADCT4565 standard; protein; 305 AA.
XX AC ADCT4565;
XX 01-JAN-2004 (first entry)
DT Human secreted protein - SEQ ID 1198.

antianemic; antirheumatic; antiarthritic; antiinflammatory; antithyroid;
antidiabetic; immunosuppressive; dermatologic; nephrotoxic;
antiparkinsonian; neuroprotective; nootropic; antibacterial; virucide;
fungicide; antiparasitic; antiarteriosclerotic; vulnery; cytostatic;
haemopoietic; haematologic; anaemia; autoimmune disorder;
rheumatoid arthritis; inflammation; Grave's disease; diabetes;
systemic lupus erythematosus; glomerulonephritis; neurodegenerative;
Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;
cancer; bacterial; viral; fungal; parasitic infection; gene therapy;
human.

OS Homo sapiens.
XX WO2003038063-A2.
XX 08-MAY-2003.
PD 19-MAR-2002; 2002WO-US008277.
XX 21-MAR-2001; 2001US-0277340P.
XX 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX MPI: 2003-430516/40.
XX N-PSDB; ADCT3950.
DR

New human secreted polypeptide for diagnosing, preventing or treating
PT hematopoietic or hematologic disorders (e.g. anemia), autoimmune
PT disorders (e.g. diabetes) or hyperproliferative disorders (e.g. cancer or
PT atherosclerosis).

XX

PS Claim 16; SEQ ID NO 1198; 2272pp; English.

CC The invention relates to a novel human secreted polypeptide comprising a
CC defined sequence given in the specification. The polypeptide, nucleic
CC acid molecule, antibody, agonist or antagonist of the invention may be
CC useful for preparing a composition for diagnosing or treating a
CC haemopoietic or haematologic disorder such as anaemia, autoimmune
CC disorders such as rheumatoid arthritis, inflammation, Grave's disease,
CC diabetes, systemic lupus erythematosus or glomerulonephritis,
CC neurodegenerative disorders including Parkinson's disease and Alzheimer's
CC disease, wounds and hyperproliferative disorders including
CC atherosclerosis or cancer, as well as bacterial, viral, fungal or
CC parasitic infections. The polypeptide may also be used during gene
CC therapy procedures and for identifying a binding partner by contacting
CC the polypeptide with a binding partner and determining whether the
CC binding partner increases or decreases the activity of the polypeptide.
CC The current sequence is that of the human secreted protein of the
CC invention.

CC Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 7; Length 305;

Best Local Similarity 100.0%; Pred. No. 3.2e-161;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLGLSGAPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLGLSGAPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
QY 61 PHLRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRRNAYSGIIGIMHEMIANNFTG 120
DB 61 PHLRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRRNAYSGIIGIMHEMIANNFTG 120
QY 121 MMRRDDACRSRSRSGKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALVYPTL 180
DB 121 MMRRDDACRSRSRSGKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALVYPTL 180
QY 181 PEALQROMQVEQDLADELITPQGHKELRLTFEDAGYLKTEENEPQLLEGSPSLGFE 240
DB 181 PEALQROMQVEQDLADELITPQGHKELRLTFEDAGYLKTEENEPQLLEGSPSLGFE 240
QY 241 TLENCRKAHKEISKEIKRLKGLLTQHGIPYTPETSNLEHGHETPRAPKSPQOLRGDPG 300
DB 241 TLENCRKAHKEISKEIKRLKGLLTQHGIPYTPETSNLEHGHETPRAPKSPQOLRGDPG 300
QY 301 LRGS 305
DB 301 LRGS 305

RESULT 11

ADD27815 ADD27815 standard; protein; 305 AA.

AC ADD27815;

DT 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase gamma subunit.

KM human; protein phosphorylation; soluble GlcNAc-phosphotransferase;

KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.

OS Homo sapiens.

XX US2003119088-A1.

XX 26-JUN-2003.

XX 21-DEC-2001; 2001US-00023888.

PR 21-DEC-2001; 2001US-00023888.

PA (NOVA-) NOVAZYME PHARM INC.

PI Canfield W, Kudo M;

DR WPI; 2003-801323/75.

XX N-PSDB; ADD27814.

PT Phosphorylating a protein for treating a patient suffering from a

PT lysosomal storage disease e.g. Fabry's disease by contacting the protein

PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated

PS Claim 11; SEQ ID NO 7; 55pp; English.

CC The invention relates to a method of phosphorylating a protein comprising
CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
CC acetylglucosamine) and producing a phosphorylated protein. The method is
CC useful for treating a patient suffering from a lysosomal storage disease
CC e.g. Fabry's disease. The present sequence represents the amino acid
CC sequence of the human GlcNAc-phosphotransferase gamma subunit.

CC Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 7; Length 305;

Best Local Similarity 100.0%; Pred. No. 3.2e-161;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLGLSGAPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLGLSGAPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
QY 61 PHLRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRRNAYSGIIGIMHEMIANNFTG 120
DB 61 PHLRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRRNAYSGIIGIMHEMIANNFTG 120
QY 121 MMRRDDACRSRSRSGKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALVYPTL 180
DB 121 MMRRDDACRSRSRSGKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALVYPTL 180
QY 181 PEALQROMQVEQDLADELITPQGHKELRLTFEDAGYLKTEENEPQLLEGSPSLGFE 240
DB 181 PEALQROMQVEQDLADELITPQGHKELRLTFEDAGYLKTEENEPQLLEGSPSLGFE 240
QY 241 TLENCRKAHKEISKEIKRLKGLLTQHGIPYTPETSNLEHGHETPRAPKSPQOLRGDPG 300
DB 241 TLENCRKAHKEISKEIKRLKGLLTQHGIPYTPETSNLEHGHETPRAPKSPQOLRGDPG 300
QY 301 LRGS 305
DB 301 LRGS 305

RESULT 12

ABM01490 ABM01490 standard; protein; 305 AA.

AC ABM01490;

DT 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase gamma subunit protein.

KM Mannoase glycoprotein; gene therapy; carboxamide deficient cell;

KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;

KM gastrointestinal; human; enzyme; lectin resistant cell;

XX deoxymannojirimycin; kifunensine; glycosylation inhibition.

XX Homo sapiens.

XX Key

FT Peptide 1. 24

FT Protein /label= Signal_peptide

FT 25. .305

FT /note= "Mature human GlcNAc-phosphotransferase gamma
FT subunit protein"
XX
PN US2003124652-A1.
XX
PD 03-JUL-2003.
XX
PF 21-DEC-2001; 2001US-00023889.
XX
PR 21-DEC-2001; 2001US-00023889.
XX
PA (NOVA-) NOVARTIS PHARM INC.
XX
PI Canfield WM;
XX
DR WPI; 2003-810984/76.
XX
NR N-PSDB; AAD62492.
XX
PT Producing a high mannose glycoprotein for treating lysosomal storage
XX disease, comprises culturing the lectin resistant mammalian cell in the
XX presence of deoxymannojirimycin and kifunensine.
XX
PS Claim 9; Page 20; 46pp; English.
XX
CC The invention relates to a method for producing a high mannose
CC glycoprotein. The method comprises: introducing and expressing a
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
CC the cell in the presence of a lectin to obtain a lectin resistant cell;
CC isolating the cell; culturing the cell in the presence of
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
CC glycoprotein; and collecting the glycoprotein. The invention is useful in
CC gene therapy. The method is useful for producing a high mannose
CC glycoprotein in a complex carbohydrate deficient cell for treating
CC lysosomal storage disease. The present sequence is human N-
CC acetylglucosamine-1 (GlcNAc)-phosphotransferase gamma subunit protein
XX
SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 7; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARIILLGLSAGPAPAGAAKMKVVEENAFGVNPPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARIILLGLSAGPAPAGAAKMKVVEENAFGVNPPFLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVESTYKYFPCPHNVTOHEQTFRRNAYSGILGIWHEWEIANTFTG 120
DB 61 PVHLFRLSGKCFSLVESTYKYFPCPHNVTOHEQTFRRNAYSGILGIWHEWEIANTFTG 120
QY 121 MMRBDGACRSRSRQSVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
DB 121 MMRBDGACRSRSRQSVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
QY 181 PEALQRMQVQEDLADDELITPQGEKILRTLFEDAGYLKTPENEPTQLBGSPDSIGFE 240
DB 181 PEALQRMQVQEDLADDELITPQGEKILRTLFEDAGYLKTPENEPTQLBGSPDSIGFE 240
QY 241 TIENCRKAKEIKETKRLKGLITTOHGIPYTRPTETSNLEHGHETPPAKSPBQLRGDPG 300
DB 241 TIENCRKAKEIKETKRLKGLITTOHGIPYTRPTETSNLEHGHETPPAKSPBQLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305

RESULT 13
ADD37973
ID ADD37973 standard; protein; 305 AA.
XX
XX ADD37973;
XX

DT 15-JAN-2004 (first entry)
XX
DE Human secreted protein #156.
XX
XX human secreted protein; Antiallergic; Antiinflammatory; Antibacterial;
XX Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic.
XX
OS Homo sapiens.
XX
PN WO200290526-A2.
XX
PD 14-NOV-2002.
XX
PF 19-MAR-2002; 2002WO-US008279.
XX
PR 21-MAR-2001; 2001US-0277340P.
XX
PR 19-JUL-2001; 2001US-0306171P.
XX
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SW;
XX
DR WPI; 2003-140218/13.
XX
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing or
PT treating allergic or asthmatic disorders, or related immediate
PT hypersensitivity disorders.
XX
PS Claim 1; SEQ ID NO 455; 1323pp; English.
XX
CC The present invention relates to an isolated polypeptide or human
CC secreted protein. The polypeptides, nucleic acid molecules, antibodies or
CC their fragments, and agonists or antagonists that bind are useful for
CC preparing a diagnostic or pharmaceutical composition for diagnosing or
CC treating allergic or asthmatic disorders. The polypeptide is also useful
CC for identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases the activity of the polypeptide. The polypeptides and nucleic
CC acid molecules are also useful for detecting, preventing, diagnosing,
CC prognosticating, treating or ameliorating inflammatory disorders
CC neoplastic diseases, wound healing and disorders of epithelial cell
CC proliferation, immune disorders, cardiovascular disorders, blood-related
CC disorders, infectious diseases, endocrine disorders, or gastrointestinal
CC disorders. The nucleic acids are also useful for chromosome
CC identification, radiation hybrid mapping or long-range restriction
CC mapping, as molecular weight markers, or as hybridization or diagnostic
CC probes. The polypeptides and antibodies are useful for providing
CC immunological probes for differential identification of the tissues
CC immunohistochemistry assays. The present sequence represents a human
CC secreted protein.
XX
SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 7; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARIILLGLSAGPAPAGAAKMKVVEENAFGVNPPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARIILLGLSAGPAPAGAAKMKVVEENAFGVNPPFLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVESTYKYFPCPHNVTOHEQTFRRNAYSGILGIWHEWEIANTFTG 120
DB 61 PVHLFRLSGKCFSLVESTYKYFPCPHNVTOHEQTFRRNAYSGILGIWHEWEIANTFTG 120
QY 121 MMRBDGACRSRSRQSVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
DB 121 MMRBDGACRSRSRQSVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
QY 181 PEALQRMQVQEDLADDELITPQGEKILRTLFEDAGYLKTPENEPTQLBGSPDSIGFE 240
DB 181 PEALQRMQVQEDLADDELITPQGEKILRTLFEDAGYLKTPENEPTQLBGSPDSIGFE 240

DB 161 PEALQRMQVQEDLADDELITPQGHKILRTLFEDAGYIKTPEENEPTOLEGGPDSLGE 240
 QY 241 TLNCRKAHKEISKEIKRLKGLITQHGIPYTRPTETSNLEHGHETPRAPKSPQQLRGDPG 300
 DB 241 TLNCRKAHKEISKEIKRLKGLITQHGIPYTRPTETSNLEHGHETPRAPKSPQQLRGDPG 300
 QY 301 LRGS 305
 DB 301 LRGS 305
 DB 301 LRGS 305
 RESULT 14
 ADD38084
 ID ADD38084 strand: protein; 305 AA.
 AC ADD38084;
 DT 15-JAN-2004 (first entry)
 XX
 DE Human secreted protein #267.
 XX
 DE human secreted protein; Antiallergic; Antiinflammatory; Antibacterial;
 KM Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic.
 XX
 OS Homo sapiens.
 XX
 PN MO200290526-A2.
 XX
 PD 14-NOV-2002.
 XX
 PF 19-MAR-2002; 2002WO-US008279.
 XX
 PR 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-140218/13.
 XX
 PT New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing or
 PT treating allergic or asthmatic disorders, or related immediate
 PT hypersensitivity disorders.
 PT
 PS Claim 1; SEQ ID NO 566; 1323bp; English.
 XX
 CC The present invention relates to an isolated polypeptide or human
 CC secreted protein. The polypeptides, nucleic acid molecules, antibodies or
 CC their fragments, and agonists or antagonists that bind are useful for
 CC preparing a diagnostic or pharmaceutical composition for diagnosing or
 CC treating allergic or asthmatic disorders. The polypeptide is also useful
 CC for identifying a binding partner by contacting the polypeptide with a
 CC binding partner, and determining whether the binding partner increases or
 CC decreases the activity of the polypeptide. The polypeptides and nucleic
 CC acid molecules are also useful for detecting, preventing, diagnosing,
 CC prognosticating, treating or ameliorating inflammatory disorders
 CC neoplastic diseases, wound healing and disorders of epithelial cell
 CC proliferation, immune disorders, cardiovascular disorders, blood-related
 CC disorders, infectious diseases, endocrine disorders, or gastrointestinal
 CC disorders. The nucleic acids are also useful for chromosome
 CC identification, radiation hybrid mapping or long-range restriction
 CC mapping, as molecular weight markers, or as hybridization or diagnostic
 CC probes. The polypeptides and antibodies are useful for providing
 CC immunological probes for differential identification of the tissues
 CC immunohistochemistry assays. The present sequence represents a human
 CC secreted protein.
 XX
 PS Sequence 305 AA;
 SO
 Query Match 100.0%; Score 1634; DB 7; Length 305;

Best Local Similarity 100.0%; Pred. No. 3.2e-161;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAGLARLLILLGLSAGPAPAGAAKVVSEPNAGVNNPFLPQASRLQAKRDPSPVG 60
 DB 1 MAAGLARLLILLGLSAGPAPAGAAKVVSEPNAGVNNPFLPQASRLQAKRDPSPVG 60
 QY 61 PVHLFRLSGKCFSLVSTYKTEFCPPHANTQHEQTFRMNAYSGILGITHHEWLIANTFTG 120
 DB 61 PVHLFRLSGKCFSLVSTYKTEFCPPHANTQHEQTFRMNAYSGILGITHHEWLIANTFTG 120
 QY 121 MMRDGDACRSRSRQSVELACGSNRLAHVSEPTCVYALTFETPLVCHPVALVYPTL 180
 DB 121 MMRDGDACRSRSRQSVELACGSNRLAHVSEPTCVYALTFETPLVCHPVALVYPTL 180
 QY 181 PEALQRMQVQEDLADDELITPQGHKILRTLFEDAGYIKTPEENEPTOLEGGPDSLGE 240
 DB 181 PEALQRMQVQEDLADDELITPQGHKILRTLFEDAGYIKTPEENEPTOLEGGPDSLGE 240
 QY 241 TLNCRKAHKEISKEIKRLKGLITQHGIPYTRPTETSNLEHGHETPRAPKSPQQLRGDPG 300
 DB 241 TLNCRKAHKEISKEIKRLKGLITQHGIPYTRPTETSNLEHGHETPRAPKSPQQLRGDPG 300
 QY 301 LRGS 305
 DB 301 LRGS 305
 RESULT 15
 AEM01539
 ID AEM01539 strand: protein; 305 AA.
 XX
 AC AEM01539;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human GlcNAc-phosphotransferase gamma subunit protein.
 XX
 KM Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
 KM N-acetylglucosamine-1-phosphotransferase; gene therapy; human; enzyme.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..24
 FT Protein /label= Signal_peptide
 FT 25..305
 FT /note= "Mature human GlcNAc-phosphotransferase gamma
 subunit protein"
 XX
 PN US2003124653-A1.
 XX
 PD 03-JUL-2003.
 XX
 PF 21-DEC-2001; 2001US-00023890.
 XX
 PR 21-DEC-2001; 2001US-00023890.
 XX
 PA (NOVA-) NOVAZYME PHARM INC.
 XX
 PI Canfield WM;
 XX
 DR WPI; 2003-810985/76.
 DR N-PSDB; AAD62651.
 XX
 PT Producing a glycoprotein with reduced complex carbohydrates by culturing
 PT the lectin resistant mammalian cell expressing the glycoprotein for
 PT treating lysosomal storage disease.
 XX
 PS Claim 9; Page 20; 46pp; English.
 XX
 CC The present invention provides a method of producing a glycoprotein
 CC having reduced complex carbohydrates by culturing the lectin resistant

CC mammalian cell expressing the glycoprotein. The method is useful for
 CC producing a glycoprotein with reduced complex carbohydrates for treating
 CC lysosomal storage disease. The present invention is also useful in gene
 CC therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-
 CC phosphotransferase gamma subunit precursor cDNA
 XX

SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 7; length 305;
 Best Local Similarity 100.0%; Pred. No. 3.2e-161;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAAGIARLLLLIGSAGPAPAGAAKMKVVEEPNAFGVNNPFLPQASRLQAKRDPSPVSG	60
DB	1	MAAGIARLLLLIGSAGPAPAGAAKMKVVEEPNAFGVNNPFLPQASRLQAKRDPSPVSG	60
QY	61	PVHLFRLSGKCFSLVESTYKYEFCFHNVTQHEQTFRNNAYSGLIGIWHMEIANTFTG	120
DB	61	PVHLFRLSGKCFSLVESTYKYEFCFHNVTQHEQTFRNNAYSGLIGIWHMEIANTFTG	120
QY	121	MMMRDGDACRSRROSQKVELACGKSNRLAHVSEPTCYALTFTPLVCPHALLVPTL	180
DB	121	MMMRDGDACRSRROSQKVELACGKSNRLAHVSEPTCYALTFTPLVCPHALLVPTL	180
QY	181	PEALORQWDQVEQDLADELITPQHEKILRTLFEDAGYLKTPENEPQTLEGGPDSLGE	240
DB	181	PEALORQWDQVEQDLADELITPQHEKILRTLFEDAGYLKTPENEPQTLEGGPDSLGE	240
QY	241	TLNCRKAHKEISKEIKELGILLTQHGIPYTRPFTETSNLEHIGHETPRAKSPDQLRGDPG	300
DB	241	TLNCRKAHKEISKEIKELGILLTQHGIPYTRPFTETSNLEHIGHETPRAKSPDQLRGDPG	300
QY	301	LRGSL 305	
DB	301	LRGSL 305	

Search completed: July 26, 2004, 11:09:12
 Job time : 30.138 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:06:34 ; Search time 8.59705 Seconds

(without alignments)
1831.549 Million cell updates/sec

Title: US-10-657-280-3
Perfect score: 1634
Sequence: 1 MAAGLARLLILGLSAGGPA.....TPRAKSPQRLGPDGRSL 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1634	100.0	305	4	US-09-635-872A-3
2	1634	100.0	305	4	US-09-636-077A-3
3	1634	100.0	305	4	US-09-636-060C-3
4	1634	100.0	305	4	US-09-986-552-3
5	1288.5	78.9	307	4	US-09-635-872A-9
6	1288.5	78.9	307	4	US-09-636-077A-9
7	1288.5	78.9	307	4	US-09-636-060C-9
8	1288.5	78.9	307	4	US-09-986-552-9
9	137.5	7.3	83	4	US-09-437-054A-16
10	119.5	7.3	83	4	US-09-437-054A-14
11	104	6.4	21	4	US-09-635-872A-28
12	104	6.4	21	4	US-09-636-077A-28
13	104	6.4	21	4	US-09-636-060C-28
14	104	6.4	21	4	US-09-986-552-28
15	93.5	5.7	378	4	US-09-199-637A-167
16	93.5	5.7	419	4	US-09-252-991A-21423
17	92.5	5.7	216	1	US-07-917-722-4
18	91.5	5.6	216	1	US-08-489-666C-5
19	91.5	5.6	216	3	US-09-057-963A-4
20	91.5	5.6	216	3	US-09-057-963A-4
21	90	5.5	2584	4	US-09-936-135-4
22	88	5.4	593	2	US-08-900-927-4
23	88	5.4	593	2	US-09-191-279-4
24	88	5.4	593	3	US-09-334-476-4
25	86	5.3	192	4	US-09-914-258-25
26	84.5	5.2	398	4	US-09-107-532A-6267
27	84.5	5.2	398	4	US-08-406-824A-8

28	83	5.1	529	4	US-09-252-991A-29757	Sequence 29757, A
29	83	5.1	635	4	US-09-489-039A-10374	Sequence 10374, A
30	82.5	5.0	711	4	US-09-489-039A-12188	Sequence 12188, A
31	81.5	5.0	360	4	US-09-540-236-2029	Sequence 2029, Ap
32	81.5	5.0	422	3	US-09-109-204-30	Sequence 30, Appl
33	81.5	5.0	422	4	US-09-490-032-30	Sequence 30, Appl
34	81.5	5.0	819	4	US-09-976-594-369	Sequence 369, App
35	81.5	5.0	1112	4	US-09-252-991A-27256	Sequence 27256, A
36	81	5.0	744	4	US-09-252-991A-15290	Sequence 15290, A
37	80.5	4.9	292	4	US-09-724-623-81	Sequence 81, Appl
38	80	4.9	249	1	US-08-450-257-57	Sequence 57, Appl
39	80	4.9	249	1	US-08-450-246-57	Sequence 57, Appl
40	80	4.9	249	1	US-08-450-098-57	Sequence 57, Appl
41	80	4.9	249	1	US-08-451-233-57	Sequence 57, Appl
42	80	4.9	249	1	US-08-450-236-57	Sequence 57, Appl
43	80	4.9	249	4	US-08-235-403-57	Sequence 57, Appl
44	80	4.9	312	1	US-08-094-128A-27	Sequence 27, Appl
45	80	4.9	312	1	US-08-455-674-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-09-635-872A-3
; Sequence 3, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OR INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
US-09-635-872A-3

Query Match Best Local Similarity 100.0%; Pred. No. 5.6e-178; Length 305;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAAGLARLLILGLSAGGPAAGAAKXVEEENAFVNNPFLPQASRLQARRDPSVSG	60
DB	1	MAAGLARLLILGLSAGGPAAGAAKXVEEENAFVNNPFLPQASRLQARRDPSVSG	60
QY	61	PVHLFRLSGKCFELVSTYKFERCPFHNTQHEQTRNMAVSGIIGIMHEWELANTFTG	120
DB	61	PVHLFRLSGKCFELVSTYKFERCPFHNTQHEQTRNMAVSGIIGIMHEWELANTFTG	120
QY	121	MMNRDDACGRSRQSGKVELACGKSNRLAHVSEPTCYATLTFETPLVCPHALVYPTL	180
DB	121	MMNRDDACGRSRQSGKVELACGKSNRLAHVSEPTCYATLTFETPLVCPHALVYPTL	180
QY	181	PEALQOMQOVVEDLDELITPGGHEKRLRTLFEDAGYKTPBENPFTLGGPDSLGRE	240
DB	181	PEALQOMQOVVEDLDELITPGGHEKRLRTLFEDAGYKTPBENPFTLGGPDSLGRE	240
QY	241	TLENCKRAHELSKEIKRLKGLITTOGIPYTRPTETSNLEHGHETPRAKSPFOALRGDGG	300
DB	241	TLENCKRAHELSKEIKRLKGLITTOGIPYTRPTETSNLEHGHETPRAKSPFOALRGDGG	300
QY	301	LRGSL 305	
DB	301	LRGSL 305	

```
RESULT 2
US-09-636-077A-3
; Sequence 3, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
US-09-636-077A-3

Query Match          100.0%; Score 1634; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 5.6e-178;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVESTYKYEFCEPHNVTQHEQTFRNMAVSGIIGIWHMEIANTFTG 120
DB 61 PVHLFRLSGKCFSLVESTYKYEFCEPHNVTQHEQTFRNMAVSGIIGIWHMEIANTFTG 120
QY 121 MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
DB 121 MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
QY 181 PEALQRQMDQVEQDLADELITPQGHKRLRTLFEDAGYIKTPENEPTQLEGGPDLSGFE 240
DB 181 PEALQRQMDQVEQDLADELITPQGHKRLRTLFEDAGYIKTPENEPTQLEGGPDLSGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPPTETSNLEHGHETPRAKSPDQLRGDPG 300
DB 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPPTETSNLEHGHETPRAKSPDQLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305

RESULT 3
US-09-636-060C-3
; Sequence 3, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US0CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION:
US-09-636-060C-3

Query Match          100.0%; Score 1634; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 5.6e-178;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVESTYKYEFCEPHNVTQHEQTFRNMAVSGIIGIWHMEIANTFTG 120
DB 61 PVHLFRLSGKCFSLVESTYKYEFCEPHNVTQHEQTFRNMAVSGIIGIWHMEIANTFTG 120
QY 121 MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
DB 121 MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
QY 181 PEALQRQMDQVEQDLADELITPQGHKRLRTLFEDAGYIKTPENEPTQLEGGPDLSGFE 240
DB 181 PEALQRQMDQVEQDLADELITPQGHKRLRTLFEDAGYIKTPENEPTQLEGGPDLSGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPPTETSNLEHGHETPRAKSPDQLRGDPG 300
DB 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPPTETSNLEHGHETPRAKSPDQLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305

RESULT 4
US-09-986-552-3
; Sequence 3, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION:
US-09-986-552-3

Query Match          100.0%; Score 1634; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 5.6e-178;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVESTYKYEFCEPHNVTQHEQTFRNMAVSGIIGIWHMEIANTFTG 120
DB 61 PVHLFRLSGKCFSLVESTYKYEFCEPHNVTQHEQTFRNMAVSGIIGIWHMEIANTFTG 120
QY 121 MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
DB 121 MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
QY 181 PEALQRQMDQVEQDLADELITPQGHKRLRTLFEDAGYIKTPENEPTQLEGGPDLSGFE 240
DB 181 PEALQRQMDQVEQDLADELITPQGHKRLRTLFEDAGYIKTPENEPTQLEGGPDLSGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPPTETSNLEHGHETPRAKSPDQLRGDPG 300
DB 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPPTETSNLEHGHETPRAKSPDQLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305
```

```

Db      121 MMRDGDACRSRSGKVELACGKSNRLAHVSEPTCYALTFETPLVCHPHALLVPTL 180
Qy      181 PEARLORQDQVEODLADDELITPOGHEKILRTLFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Db      181 PEARLORQDQVEODLADDELITPOGHEKILRTLFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Qy      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPPAKSPEDLRGDPG 300
Db      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPPAKSPEDLRGDPG 300
Qy      301 LRGS 305
Db      301 LRGS 305

RESULT 5
US-09-635-872A-9
; Sequence 9, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 9
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-635-872A-9.

```

Query Match 78.9%; Score 1288.5; DB 4; Length 307;

Best Local Similarity 78.4%; Pred. No. 1.7e-138;

Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

```

Qy      1 MAAGLARLLILLGLSAGPAPAGAAKMKVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
Db      1 MAGRLAGFLMLLGLASQGPAPACAGMKVVEEPNTFGLNPFLLPQASRLQKREBSAVSG 60
Qy      61 PVLRLSGKCFSLVSTYKYECPPHANTQHEOTFRMNAYSGLIGIWHMEIANNFTG 120
Db      61 PVLRLSGKCFSLVSTYKYECPPHANTQHEOTFRMNAYSGLIGIWHMEIANNFTG 120
Qy      121 MMRDGDACRSRSGKVELACGKSNRLAHVSEPTCYALTFETPLVCHPHALLVPTL 180
Db      121 MMRDGDACRSRSGKVELACGKSNRLAHVSEPTCYALTFETPLVCHPHALLVPTL 180
Qy      181 PEARLORQDQVEODLADDELITPOGHEKILRTLFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Db      181 PEARLORQDQVEODLADDELITPOGHEKILRTLFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Qy      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPPAKSPEDLRGDP 299
Db      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPPAKSPEDLRGDP 299
Qy      300 LRGS 305
Db      301 LRGS 306

RESULT 6
US-09-636-077A-9
; Sequence 9, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0

```

```

; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 9
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-077A-9

```

Query Match 78.9%; Score 1288.5; DB 4; Length 307;

Best Local Similarity 78.4%; Pred. No. 1.7e-138;

Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

```

Qy      1 MAAGLARLLILLGLSAGPAPAGAAKMKVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
Db      1 MAGRLAGFLMLLGLASQGPAPACAGMKVVEEPNTFGLNPFLLPQASRLQKREBSAVSG 60
Qy      61 PVLRLSGKCFSLVSTYKYECPPHANTQHEOTFRMNAYSGLIGIWHMEIANNFTG 120
Db      61 PVLRLSGKCFSLVSTYKYECPPHANTQHEOTFRMNAYSGLIGIWHMEIANNFTG 120
Qy      121 MMRDGDACRSRSGKVELACGKSNRLAHVSEPTCYALTFETPLVCHPHALLVPTL 180
Db      121 MMRDGDACRSRSGKVELACGKSNRLAHVSEPTCYALTFETPLVCHPHALLVPTL 180
Qy      181 PEARLORQDQVEODLADDELITPOGHEKILRTLFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Db      181 PEARLORQDQVEODLADDELITPOGHEKILRTLFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Qy      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPPAKSPEDLRGDP 299
Db      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPPAKSPEDLRGDP 299
Qy      300 LRGS 305
Db      301 LRGS 306

```

RESULT 7

US-09-636-060C-9

```

; Sequence 9, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US0CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 9
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-060C-9

```

Query Match 78.9%; Score 1288.5; DB 4; Length 307;

Best Local Similarity 78.4%; Pred. No. 1.7e-138;

Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

```

Qy      1 MAAGLARLLILLGLSAGPAPAGAAKMKVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
Db      1 MAGRLAGFLMLLGLASQGPAPACAGMKVVEEPNTFGLNPFLLPQASRLQKREBSAVSG 60
Qy      61 PVLRLSGKCFSLVSTYKYECPPHANTQHEOTFRMNAYSGLIGIWHMEIANNFTG 120
Db      61 PVLRLSGKCFSLVSTYKYECPPHANTQHEOTFRMNAYSGLIGIWHMEIANNFTG 120

```

QY 121 MMARDGACRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
DB 121 MMARDGACRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
QY 181 PEALORQMDQVEODLADDELITPQGEKILRTLFEDAGYLKPEENEPQLBGGPDSLGE 240
DB 181 SEALQORLDQVEODLADDELITPQGEKILRTLFEDAGYLKPEENEPQLBGGPDSLGE 240
QY 241 TLENCKRAHAEISQEVQRLTSLQOHGIPHTPTETSHQHLGQQLFGALIAAEHLRSDP 299
DB 241 TLENCKRAHAEISQEVQRLTSLQOHGIPHTPTETSHQHLGQQLFGALIAAEHLRSDP 300
QY 300 GLRGSL 305
DB 301 GLRGNI 306

RESULT 8
US-09-986-552-9
; Sequence 9, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 21508905/7DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-986-352-9

Query Match 78.9%; Score 1288.5; DB 4; Length 307;
Best Local Similarity 78.4%; Pred. No. 1.7e-138;
Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;
QY 1 MAGIARLLLLGLSAGGAPAGAAKMKVVEEPNAFGVNPFPLPQASLQAKRDPSPYSG 60
DB 1 MGRLAGFLMLGLASQGPAPACAKMKVVEEPNFGLNPFPLPQASLQAKRDPSPYSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYFCFPHVNTCHEQTFRNAYSGILGIWHEMIANNFTG 120
DB 61 PHLFRLAGKCFSLVSTYKYFCFPHVNTCHEQTFRNAYSGILGIWHEMIANNFTG 120
QY 121 MMARDGACRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
DB 121 MMARDGACRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
QY 181 PEALORQMDQVEODLADDELITPQGEKILRTLFEDAGYLKPEENEPQLBGGPDSLGE 240
DB 181 SEALQORLDQVEODLADDELITPQGEKILRTLFEDAGYLKPEENEPQLBGGPDSLGE 240
QY 241 TLENCKRAHAEISQEVQRLTSLQOHGIPHTPTETSHQHLGQQLFGALIAAEHLRSDP 299
DB 241 TLENCKRAHAEISQEVQRLTSLQOHGIPHTPTETSHQHLGQQLFGALIAAEHLRSDP 300
QY 300 GLRGSL 305
DB 301 GLRGNI 306

RESULT 9
US-09-437-054A-16
; Sequence 16, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:

; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: B81273 US NA
; CURRENT APPLICATION NUMBER: US/09/437,054A
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/107,909
; PRIOR FILING DATE: 1998-No. 6316698member-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-437-054A-16

Query Match 8.4%; Score 137.5; DB 4; Length 109;
Best Local Similarity 31.7%; Pred. No. 1e-07;
Matches 40; Conservative 13; Mismatches 50; Indels 23; Gaps 5;
QY 84 CPFHANTQHEQTFRNAYSGILGIWHEMIANNFTGMMARDGAC-RSRSGKVELAC 142
DB 6 CPYKASQVEGHSSTN-----LGRMDKE---ESCRMHFSNGDKCMNGPDSILKRLRC 57
QY 143 GKSNNLAHVSEPTCYVALTFETPLVCHPHALLVPTLPEALORQMDQVEODLADDELITP 202
DB 58 GLSNEINGVDEPSCRYAVLSTPAMCVBEKL-----KELQOKLDAASDLS----- 104
QY 203 QGHEKL 208
DB 105 -GHDEL 109

RESULT 10
US-09-437-054A-14
; Sequence 14, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: B81273 US NA
; CURRENT APPLICATION NUMBER: US/09/437,054A
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/107,909
; PRIOR FILING DATE: 1998-No. 6316698member-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-437-054A-14

Query Match 7.3%; Score 119.5; DB 4; Length 83;
Best Local Similarity 35.6%; Pred. No. 7.5e-06;
Matches 31; Conservative 7; Mismatches 40; Indels 9; Gaps 3;
QY 84 CPFHANTQHEQTFRNAYSGILGIWHEMIANNFTGMMARDGAC-RSRSGKVELAC 142
DB 3 CPYKASQVEGHSSTN-----LGRMDKE---ESCRMHFSNGDKCMNGPDSILKRLRC 54
QY 143 GKSNNLAHVSEPTCYVALTFETPLVC 169
DB 55 GLSNEINGVDEPSCRYAVLSTPAMC 81

RESULT 11
US-09-635-872A-28
; Sequence 28, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM

```

?
? TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
?
? FILE REFERENCE: 1956131US0
?
? CURRENT APPLICATION NUMBER: US/09/635,872A
?
? CURRENT FILING DATE: 2000-08-10
?
? PRIOR APPLICATION NUMBER: 60/153,831
?
? PRIOR FILING DATE: 1999-09-14
?
? NUMBER OF SEQ ID NOS: 52
?
? SOFTWARE: PatentIn version 3.0
?
? SEQ ID NO 28
?
? LENGTH: 21
?
? TYPE: PR1
?
? ORGANISM: Bos taurus
?
US-09-635-872A-28

```

Query Match	6.4%	Score 104;	DB 4;	Length 21;
Best Local Similarity	90.5%;	Pred. No. 4.8e-05;		
Matches 19; Conservative	1;	Mismatches 1;	Indels 0;	Gaps 0

QY 25 AKMKVVEEPNAFGVNNPFLPQ 45
||| ||| : ||| |||
Db 1 AKMKVVEEPNTFGLNPFLLPQ 21

```

RESULT 12
US-09-636-077A-28
: Sequence 28, Application US/09636077A
: Patent No. 6537785
: GENERAL INFORMATION:
: APPLICANT: CANFIELD, WILLIAM
: TITLE OF INVENTION: METHODS OF TREATING LYSSOMAL STORAGE DISEASES
: FILE REFERENCE: 195612U90
: CURRENT APPLICATION NUMBER: US/09/636,077A
: CURRENT FILING DATE: 2000-08-10
: PRIOR APPLICATION NUMBER: 60/153,831
: PRIOR FILING DATE: 1999-09-14
: NUMBER OF SEQ ID NOS: 52
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 28
: LENGTH: 21
: TYPE: PR1
: ORGANISM: Bos taurus
: US-09-636-077A-28

```

Query Match	6.4%	Score 104;	DB 4;	Length 21;
Best Local Similarity	90.5%;	Pred. No. 4.8e-05;		
Matches 19; Conservative	1;	Mismatches 1;	Indels 0;	Gaps 0;

QY 25 AKMKVVEEENAFGVNPFLLPQ 45
|||:|||||
Db 1 AKMKVVEEENTFGLNPFLLPQ 21

```

RESULT 13
US-09-636-060C-28
; Sequence 28, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US00CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Version 3.1
; SEQ ID NO 28
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Bos taurus
; US-09-636-060C-28

```

Query Match 6.4%; Score 104; DB 4; Length 21;

Best Local Similarity 90.5%; Pred. No. 4.8e-05;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 25 AKMKVVEEPNAFGVNNPFLPQ 45
|||||:|||||
Db 1 AKMKVVEEPNFTGLANNPFLPQ 21

QY 25 AKMKVVEEPNAGVGNPFLPQ 45
||||| : |||||
Db 1 AKMKVVEEPNTFGLNPFLLPQ 21

```

RESULT 14
US-09-986-552-28
; Sequence 28, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215085US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 21
; TYPE: prt
; ORGANISM: Bos taurus
US-09-986-552-28

```

Query Match	6.4%	Score 104	DB 4	length 21
Best Local Similarity	90.5%	Pred. No.	4.8e-05	
Matches	19	Conservative	1	Mismatches 1
				Indels 0
				Gaps 0

```
QY      25 AKMKVVEEPPNAFGVNNPFLPQ 45
          |||||:|||||
Db      1 AKMKVVEEPPNTFGLNPPFLPQ 21
```

```

RESULT 15
; Sequence 367, Application US/0919637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Wikios, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drinkard, Bliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: 'VIOLENCE ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 367
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-199-637A-367

```

Query March	5.7%	Score	93.5	DB	4	Length	378
Best Local Similarity	20.8%	Pred. No.	0.082				
Matches	57	Conservative	28	Mismatches	76	Indels	113
						Gaps	11

23 GAAGKMKVVEEFNFAFGVANNPFLPQASBLQAKRDPVSGPVHLFRLSGKCFSLVESTKYK 82


```

Db      145 GAGFORAVYMSGVAVNRPRLARSLRASIQQ----- 175
Qy      83 FCFPHAVTQHEOTFRMNNAYSGILGIWHEMEIANTFTGMWEDGD-----ACRSRSRQS 136
Db      176 ---FANLELHEQT-----EVRG-----WLRDGRVVGVAATSRGEIRGD 210
Qy      137 KYELACGK-SNRLAHVSEPTCVALTETPTVCHPHALLVYPTLPEALQRQMDQVEODL 195
Db      211 KTLAAGAMSGEL-----LKPLGJELPVVPVKGOMILYKCAADFLPR----- 252
Qy      196 ADELITPQCH-----EKLIRTLPEDAGYLKTPEBNEPTQLBGCPDLSLGFETLNCRK 247
Db      253 ---MYLAKGRVAIPRRDGHILIGSTLEHSGFDKPTD-----EALLESIRA 294
Qy      248 AHKELSKXIKRLK-----GLL--TOHGIPYTRP 273
Db      295 SNAELIPELADMQPVAHMAGLRPGSPGIFPIGP 328

```

Search completed: July 26, 2004, 11:13:20
 Job time : 9.59705 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:11:44 ; Search time 23.4465 Seconds
(without alignments)
4073.614 Million cell updates/sec

Title:	US-10-657-280-3
Perfect score:	1634
Sequence:	1 MAAGLARLLLLIGLSAGPA.....TPRAKSPEQLRGDPGLRGSL 309

Scoring table: BLOSUM62

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpa/PCr_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpa/PCrTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpa/US09_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubpa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubpa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubpa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubpa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubpa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubpa/US10_NEW_PUB.pcp.*
17: /cgn2_6/ptodata/1/pubpa/US06_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/1/pubpa/US06_PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1634	100.0	305	9	US-09-895-072-3	Sequence 3, Appli
2	1634	100.0	305	9	US-09-886-552-3	Sequence 3, Appli
3	1634	100.0	305	14	US-10-023-888-7	Sequence 7, Appli
4	1634	100.0	305	14	US-10-023-888-7	Sequence 7, Appli
5	1634	100.0	305	14	US-10-023-889-7	Sequence 7, Appli
6	1634	100.0	305	14	US-10-023-890-7	Sequence 7, Appli
7	1634	100.0	305	14	US-10-023-894-7	Sequence 7, Appli
8	1634	100.0	305	14	US-10-306-686-3	Sequence 3, Appli
9	1634	100.0	306	12	US-10-327-876-152	Sequence 152, Appli
10	1634	100.0	306	14	US-10-037-085-15	Sequence 15, Appli
11	1634	100.0	307	12	US-10-296-115-1319	Sequence 1319, Appli
12	1288.5	78.9	307	9	US-09-895-072-9	Sequence 9, Appli
13	1288.5	78.9	307	9	US-09-886-552-9	Sequence 9, Appli
14	1288.5	78.9	307	14	US-10-023-888-12	Sequence 12, Appli
15	1288.5	78.9	307	14	US-10-023-889-12	Sequence 12, Appli

16	1288.5	78.9	307	14	US-10-023-690-12	Sequence 12, App
17	1288.5	78.9	307	14	US-10-024-197-12	Sequence 12, App
18	1288.5	78.9	307	14	US-10-023-894-12	Sequence 12, App
19	1288.5	78.9	307	14	US-10-306-686-9	Sequence 12, App
20	1244.5	76.2	296	15	US-10-131-687-112	Sequence 11, App
21	237	14.5	38	14	US-10-029-686-31490	Sequence 31490, App
22	205	12.5	34	14	US-10-029-865-31804	Sequence 31804, App
23	170	10.4	478	12	US-10-424-539-276439	Sequence 276439, App
24	169.5	10.4	432	12	US-10-425-114-47678	Sequence 47678, App
25	167.5	10.3	490	12	US-10-424-599-276502	Sequence 276502, App
26	156.5	9.6	678	12	US-10-425-114-41458	Sequence 41458, App
27	156.5	9.6	610	12	US-10-425-114-453637	Sequence 453637, App
28	155.5	9.5	597	15	US-10-331-6964-39	Sequence 39, App
29	155.5	9.5	612	12	US-10-072-012-650	Sequence 650, App
30	155.5	9.5	650	12	US-10-072-012-244	Sequence 244, App
31	155.5	9.5	667	12	US-10-072-012-649	Sequence 649, App
32	154	9.4	614	16	US-10-337-663-111038	Sequence 111038, App
33	149	9.1	668	12	US-10-072-012-242	Sequence 242, App
34	140.5	8.6	524	12	US-10-221-278-212	Sequence 212, App
35	140.5	8.6	524	15	US-10-291-172-212	Sequence 212, App
36	140.5	8.6	527	12	US-10-221-278-211	Sequence 211, App
37	140.5	8.6	527	12	US-10-257-021-220	Sequence 20, App
38	140.5	8.6	527	15	US-10-391-172-211	Sequence 211, App
39	140.5	8.6	527	16	US-10-408-165-484	Sequence 484, App
40	140.5	8.6	534	12	US-10-221-278-213	Sequence 213, App
41	140.5	8.6	534	15	US-10-291-172-213	Sequence 213, App
42	130	8.0	571	12	US-10-072-012-652	Sequence 652, App
43	125	7.6	483	16	US-10-408-165-1836	Sequence 1836, App
44	120.5	7.4	483	16	US-09-796-753-44	Sequence 44, App
45	119	7.3	483	12	US-10-221-278-310	Sequence 310, App

ALIGNMENTS

```

RESULT 1
US-09-895-072-3
; Sequence 3, Application US/03895072
; Patent No. US20020025550A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 210119USOCONT
; CURRENT APPLICATION NUMBER: US/09/895, 072
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/153, 831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/635, 872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION:
; US-09-895-072-3

```

Query Match	100.0%;	Score 1634;	DB 9;	Length 305;
Best Local Similarity	100.0%;	Pred. No. 1.3e-156;		
Matches 305; Conservative	0;	Mismatches	0;	Gaps .0

QY	1	MAAGGARILLILLIGLSAGSGAPAPGAAMKMYEPPNAFGVNNPFLQASLQAKRPPSPVG	60
Db	1	MAAGGARILLILLIGLSAGSGAPAPGAAMKMYEPPNAFGVNNPFLQASLQAKRPPSPVG	60
QY	61	PVHLFRLSKGCSLVESTYKVEFCFPPHNVTOBQTPRNNAASGILGIWHEMIANNPTFG	120
Db	61	PVHLFRLSKGCSLVESTYKVEFCFPPHNVTOBQTPRNNAASGILGIWHEMIANNPTFG	120


```

QY 61 PVHLFRLSGKCFSLVSTYKYEFCPFHNVTQHEQTFRNNAYSGILGIWHEWEIANNPTFG 120
DB 61 PVHLFRLSGKCFSLVSTYKYEFCPFHNVTQHEQTFRNNAYSGILGIWHEWEIANNPTFG 120
QY 121 MMNRDGDACRSRSRQSKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTL 180
DB 121 MMNRDGDACRSRSRQSKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTL 180
QY 181 PEALORQMDQVQEDLADDELITPQGEKILRTLFEADAGYLKTPBENEPQLGEGPDSLGE 240
DB 181 PEALORQMDQVQEDLADDELITPQGEKILRTLFEADAGYLKTPBENEPQLGEGPDSLGE 240
QY 241 TLENCRKAHKELSKEIKLKGILLTOHGIPYTRPETSNIJLHGHETPRAKSPQOLRGDPG 300
DB 241 TLENCRKAHKELSKEIKLKGILLTOHGIPYTRPETSNIJLHGHETPRAKSPQOLRGDPG 300
QY 301 LRGS 305
DB 301 LRGS 305

RESULT 5
US-10-023-890-7
; Sequence 7, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYD
; FILE REFERENCE: 20310US77
; CURRENT APPLICATION NUMBER: US/10/023, 890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-890-7

Query Match      100.0%; Score 1634; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS
; FILE REFERENCE: 209794US0
; CURRENT APPLICATION NUMBER: US/10/024, 197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-024-197-7

Query Match      100.0%; Score 1634; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLILLIGLSAGGPAPAGAAKMYVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLILLIGLSAGGPAPAGAAKMYVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYEFCPFHNVTQHEQTFRNNAYSGILGIWHEWEIANNPTFG 120
DB 61 PVHLFRLSGKCFSLVSTYKYEFCPFHNVTQHEQTFRNNAYSGILGIWHEWEIANNPTFG 120
QY 121 MMNRDGDACRSRSRQSKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTL 180
DB 121 MMNRDGDACRSRSRQSKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTL 180
QY 181 PEALORQMDQVQEDLADDELITPQGEKILRTLFEADAGYLKTPBENEPQLGEGPDSLGE 240
DB 181 PEALORQMDQVQEDLADDELITPQGEKILRTLFEADAGYLKTPBENEPQLGEGPDSLGE 240
QY 241 TLENCRKAHKELSKEIKLKGILLTOHGIPYTRPETSNIJLHGHETPRAKSPQOLRGDPG 300
DB 241 TLENCRKAHKELSKEIKLKGILLTOHGIPYTRPETSNIJLHGHETPRAKSPQOLRGDPG 300
QY 301 LRGS 305
DB 301 LRGS 305

RESULT 7
US-10-023-894-7
; Sequence 7, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; APPLICANT: KORNFIELD, Stuart
; TITLE OF INVENTION: EXPRESSION OF LYOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; FILE REFERENCE: 217139US77
; CURRENT APPLICATION NUMBER: US/10/023, 894
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-894-7

Query Match      100.0%; Score 1634; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLILLIGLSAGGPAPAGAAKMYVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLILLIGLSAGGPAPAGAAKMYVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYEFCPFHNVTQHEQTFRNNAYSGILGIWHEWEIANNPTFG 120
DB 61 PVHLFRLSGKCFSLVSTYKYEFCPFHNVTQHEQTFRNNAYSGILGIWHEWEIANNPTFG 120

```

Db 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNTQHEQTFRNNAYSGIIGIWHMEIANNTFTG 120
QY 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVPTL 180
Db 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVPTL 180
QY 181 PEALOROMQVQVEODLADDELITPOGHEKILRTLFEAGYLTKEEENEPQLLEGSPDSIGFE 240
Db 181 PEALOROMQVQVEODLADDELITPOGHEKILRTLFEAGYLTKEEENEPQLLEGSPDSIGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTSTNLEHGHETPRAXSPBQLRGDPG 300
Db 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTSTNLEHGHETPRAXSPBQLRGDPG 300
QY 301 LRGS 305
Db 301 LRGS 305

RESULT 8
US-10-306-686-3
; Sequence 3, Application US/10306686
; Publication No. US20030148460A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANase OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 230397US7Ddiv
; CURRENT APPLICATION NUMBER: US/10/306,686
; CURRENT FILING DATE: 2002-11-29
; PRIOR APPLICATION NUMBER: 09/636,596
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-08-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION:
US-10-306-686-3

Query Match 100.0%; Score 1634; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1,3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLILLGLSGAPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
Db 1 MAAGLARLLILLGLSGAPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNTQHEQTFRNNAYSGIIGIWHMEIANNTFTG 120
Db 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNTQHEQTFRNNAYSGIIGIWHMEIANNTFTG 120
QY 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVPTL 180
Db 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVPTL 180
QY 181 PEALOROMQVQVEODLADDELITPOGHEKILRTLFEAGYLTKEEENEPQLLEGSPDSIGFE 240
Db 181 PEALOROMQVQVEODLADDELITPOGHEKILRTLFEAGYLTKEEENEPQLLEGSPDSIGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTSTNLEHGHETPRAXSPBQLRGDPG 300
Db 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTSTNLEHGHETPRAXSPBQLRGDPG 300
QY 301 LRGS 305
Db 301 LRGS 305

RESULT 9
US-10-372-876-152
; Sequence 152, Application US/10372876
; Publication No. US20030204071A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A. et al.
; TITLE OF INVENTION: 110 Human Secreted Proteins
; FILE REFERENCE: P2021P1
; CURRENT APPLICATION NUMBER: US/10/372,876
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 09/334,595
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: PCT/US98/27059
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/070,923
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,007
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,057
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,006
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,369
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,367
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,169
; PRIOR FILING DATE: 1997-12-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 152
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (306)
; OTHER INFORMATION: Xaa equals stop translation
US-10-372-876-152

Query Match 100.0%; Score 1634; DB 12; Length 306;
Best Local Similarity 100.0%; Pred. No. 1,3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLILLGLSGAPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
Db 1 MAAGLARLLILLGLSGAPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNTQHEQTFRNNAYSGIIGIWHMEIANNTFTG 120
Db 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNTQHEQTFRNNAYSGIIGIWHMEIANNTFTG 120
QY 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVPTL 180
Db 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVPTL 180
QY 181 PEALOROMQVQVEODLADDELITPOGHEKILRTLFEAGYLTKEEENEPQLLEGSPDSIGFE 240
Db 181 PEALOROMQVQVEODLADDELITPOGHEKILRTLFEAGYLTKEEENEPQLLEGSPDSIGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTSTNLEHGHETPRAXSPBQLRGDPG 300
Db 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTSTNLEHGHETPRAXSPBQLRGDPG 300
QY 301 LRGS 305
Db 301 LRGS 305

```
RESULT 10
US-10-097-065-152
; Sequence 152, Application US/10097065
; Publication No. US20030055236A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A. et al.
; TITLE OF INVENTION: 110 Human Secreted Proteins
; FILE REFERENCE: P2021PI
; CURRENT APPLICATION NUMBER: US/10/097,065
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: PCT/US98/27059
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/070,923
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,007
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,057
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,006
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,369
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,367
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,169
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,053
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,064
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,054
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,008
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,365
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 152
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (306)
; OTHER INFORMATION: Xaa equals stop translation
US-10-097-065-152

Query Match      100.0%; Score 1634; DB 14; Length 306;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLILLGLSAGGPAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLILLGLSAGGPAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYEFCPPHNVYTOHEQTRNNAYSGILGIWHEWEIANNFTFG 120
DB 61 PVHLFRLSGKCFSLVSTYKYEFCPPHNVYTOHEQTRNNAYSGILGIWHEWEIANNFTFG 120
QY 121 MMRBDGACRSRSGKVELACGSKNRLAHVSEPSCTCYALTETETLVCHPHALLVYPTL 180
DB 121 MMRBDGACRSRSGKVELACGSKNRLAHVSEPSCTCYALTETETLVCHPHALLVYPTL 180
QY 181 PEALORQWQVQDVLADDELITPQGHKLLRTLTFEDAGYLKTPENEPPTQLGGPDSLGF 240
DB 181 PEALORQWQVQDVLADDELITPQGHKLLRTLTFEDAGYLKTPENEPPTQLGGPDSLGF 240
QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPRAKSPBQLRGDPG 300
DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPRAKSPBQLRGDPG 300
```

```
DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPRAKSPBQLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305

RESULT 11
US-10-296-115-1319
; Sequence 1319, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hysq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1319
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-296-115-1319

Query Match      100.0%; Score 1634; DB 12; Length 310;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLILLGLSAGGPAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
DB 6 MAAGLARLLILLGLSAGGPAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 65
QY 61 PVHLFRLSGKCFSLVSTYKYEFCPPHNVYTOHEQTRNNAYSGILGIWHEWEIANNFTFG 120
DB 66 PVHLFRLSGKCFSLVSTYKYEFCPPHNVYTOHEQTRNNAYSGILGIWHEWEIANNFTFG 125
QY 121 MMRBDGACRSRSGKVELACGSKNRLAHVSEPSCTCYALTETETLVCHPHALLVYPTL 180
DB 126 MMRBDGACRSRSGKVELACGSKNRLAHVSEPSCTCYALTETETLVCHPHALLVYPTL 185
QY 181 PEALORQWQVQDVLADDELITPQGHKLLRTLTFEDAGYLKTPENEPPTQLGGPDSLGF 240
DB 186 PEALORQWQVQDVLADDELITPQGHKLLRTLTFEDAGYLKTPENEPPTQLGGPDSLGF 245
QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPRAKSPBQLRGDPG 300
DB 246 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPRAKSPBQLRGDPG 305
QY 301 LRGSLL 305
DB 306 LRGSLL 310

RESULT 12
US-09-895-072-9
; Sequence 9, Application US/09895072
; Patent No. US2002002550A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 210119US00CONT
; CURRENT APPLICATION NUMBER: US/09/895,072
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/635,872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent In version 3.1
```

SEQ ID NO 9
 LENGTH: 307
 TYPE: PRP
 ORGANISM: Mus musculus
 US-09-895-072-9

Query Match 78.9%; Score 1288.5; DB 9; Length 307;
 Best Local Similarity 78.4%; Pred. No. 1.3e-121;
 Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

QY 1 MAAGLARIILLLGLSAGAPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPYSG 60
 DB 1 MAGRLAGFLMLGLSAGAPAPAGAAKMKVVEEPNFGNNPFLPQASRLQAKRDPSPYSG 60
 QY 61 PVLHRLSGKCFSLVSTYKVFCEPHNVTQHEQTFRNNAVSGIIGIMHEMIANNFTFG 120
 DB 61 PVLHRLSGKCFSLVSTYKVFCEPHNVTQHEQTFRNNAVSGIIGIMHEMIANNFTFG 120
 QY 121 MMRRDDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALVYPTL 180
 DB 121 MMRTDGSCHSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALVYPTL 180
 QY 181 PEALORQWQVEQDLADELITPQGHKELRLTFEDAGYLYKTPNEENPTQLSGPDSIGFE 240
 DB 181 SEALQRLQVQDLADELITPQGHKELRLTFEDAGYLYKTPNEENPTQLSGPDSIGFE 240
 QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETP-RAKSPQRLGDP 299
 DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETP-RAKSPQRLGDP 299
 QY 300 GLRGSL 305
 DB 301 GLRGNI 306

RESULT 13

US-09-986-552-9
 Sequence 9; Application US/09986552
 Patent No. US20020150981A1
 GENERAL INFORMATION:
 APPLICANT: CANFIELD, William
 TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
 FILE REFERENCE: 215089US77DIV
 CURRENT APPLICATION NUMBER: US/09/986,552
 CURRENT FILING DATE: 2001-11-09
 PRIOR FILING DATE: 2000-08-10
 PRIOR APPLICATION NUMBER: 09/635,872
 PRIOR FILING DATE: 2000-08-10
 PRIOR APPLICATION NUMBER: 60/153,831
 PRIOR FILING DATE: 1999-09-14
 NUMBER OF SEQ ID NOS: 52
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 9
 LENGTH: 307
 TYPE: PRP
 ORGANISM: Mus musculus
 US-09-986-552-9

Query Match 78.9%; Score 1288.5; DB 9; Length 307;
 Best Local Similarity 78.4%; Pred. No. 1.3e-121;
 Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

QY 1 MAAGLARIILLLGLSAGAPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPYSG 60
 DB 1 MAGRLAGFLMLGLSAGAPAPAGAAKMKVVEEPNFGNNPFLPQASRLQAKRDPSPYSG 60
 QY 61 PVLHRLSGKCFSLVSTYKVFCEPHNVTQHEQTFRNNAVSGIIGIMHEMIANNFTFG 120
 DB 61 PVLHRLSGKCFSLVSTYKVFCEPHNVTQHEQTFRNNAVSGIIGIMHEMIANNFTFG 120
 QY 121 MMRRDDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALVYPTL 180
 DB 121 MMRTDGSCHSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALVYPTL 180

QY 181 PEALORQWQVEQDLADELITPQGHKELRLTFEDAGYLYKTPNEENPTQLSGPDSIGFE 240
 DB 181 SEALQRLQVQDLADELITPQGHKELRLTFEDAGYLYKTPNEENPTQLSGPDSIGFE 240
 QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETP-RAKSPQRLGDP 299
 DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETP-RAKSPQRLGDP 299
 QY 300 GLRGSL 305
 DB 301 GLRGNI 306

RESULT 14

US-10-023-888-12
 Sequence 12; Application US/10023888
 Publication No. US20030119068A1
 GENERAL INFORMATION:
 APPLICANT: CANFIELD, William
 TITLE OF INVENTION: SOLUBLE GLCNAc PHOSPHOTRANSFERASE
 FILE REFERENCE: 203515US77
 CURRENT APPLICATION NUMBER: US/10/023,888
 CURRENT FILING DATE: 2001-12-21
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 12
 LENGTH: 307
 TYPE: PRP
 ORGANISM: Mus musculus
 US-10-023-888-12

Query Match 78.9%; Score 1288.5; DB 14; Length 307;
 Best Local Similarity 78.4%; Pred. No. 1.3e-121;
 Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

QY 1 MAAGLARIILLLGLSAGAPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPYSG 60
 DB 1 MAGRLAGFLMLGLSAGAPAPAGAAKMKVVEEPNFGNNPFLPQASRLQAKRDPSPYSG 60
 QY 61 PVLHRLSGKCFSLVSTYKVFCEPHNVTQHEQTFRNNAVSGIIGIMHEMIANNFTFG 120
 DB 61 PVLHRLSGKCFSLVSTYKVFCEPHNVTQHEQTFRNNAVSGIIGIMHEMIANNFTFG 120
 QY 121 MMRRDDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALVYPTL 180
 DB 121 MMRTDGSCHSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALVYPTL 180
 QY 181 PEALORQWQVEQDLADELITPQGHKELRLTFEDAGYLYKTPNEENPTQLSGPDSIGFE 240
 DB 181 SEALQRLQVQDLADELITPQGHKELRLTFEDAGYLYKTPNEENPTQLSGPDSIGFE 240
 QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETP-RAKSPQRLGDP 299
 DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETP-RAKSPQRLGDP 299
 QY 300 GLRGSL 305
 DB 301 GLRGNI 306

RESULT 15

US-10-023-889-12
 Sequence 12; Application US/10023889
 Publication No. US20030124652A1
 GENERAL INFORMATION:
 APPLICANT: CANFIELD, William
 TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBO
 FILE REFERENCE: 203512US77
 CURRENT APPLICATION NUMBER: US/10/023,889
 CURRENT FILING DATE: 2001-12-21
 NUMBER OF SEQ ID NOS: 21
 SOFTWARE: PatentIn version 3.1

; SEQ ID NO 12
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-889-12

Query Match 78.9%; Score 1268.5; DB 14; Length 307;
Best Local Similarity 78.4%; Pred. No. 13e-12;
Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

```

QY      1 MAAGLARIILLGLSAGGPAPAPAGAAKMKVVEEPNAPGVNPPFLPQASRLQAKRDPSPVSG 60
      1 MAGRLAGFLMLGLASGGPAPACAGKMKVVEEPNTEGNNPFLPQASRLQKREPSAVSG 60
DB      61 PVHLFRLSGKCFSLVESTYKYKFCFPHNVTOHEQTFRNNAYSGLIGIWHMEIANNTFTG 120
      61 PLHLFRLAGKCFSLVESTYKYKFCFPHNVTOHEQTFRNNAYSGLIGIWHMEIINNTFKG 120
QY      121 MMRDGDACRSRSRQSKTEIACGKSNRLAHVSEPTCYALTFFETPLVCHPHALLVYPTL 180
DB      121 MMTDGDSCHSRSRQSKTEIACGKSNRLAHVSEPTCYALTFFETPLVCHPHALLVYPTL 180
QY      181 PEALQOWDQVEODIADSEITPQGEKILRTLFEDAGYLKTPENEPPTOLEGSPDSLGE 240
DB      181 SEALQORLDQVEODIADSEITPQGEKILRTLFEDAGYLKTPENEPPTOLEGSPDSLGE 240
QY      241 TLNCRKXAKHLSKEIKRLKGLLTQHGIPYTRPETSNTLEHGHETP-RAKSPQLRGDP 299
DB      241 TLNCRKXAKHLSQEVORLTSILOQHGIPYTOPETHTSHQHLGOOLPIGATAAHLRSDP 300
QY      300 GLRGSIL 305
DB      301 GLRGNI 306

```

Search completed: July 26, 2004, 11:25:26
Job time : 23.4465 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 26, 2004, 11:05:38 ; Search time 7.22934 Seconds

(without alignments)
4058.242 Million cell updates/sec

Title: US-10-657-280-3
Perfect score: 1634
Sequence: 1 MAAGLARLLILGLISAGGPA.....TPRAKSPQRLGDPGLRSL 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1634	100.0	305	2	T45062	hypothetical prote
2	185.5	11.4	507	2	T24944	hypothetical prote
3	155.5	9.5	667	2	JC5889	OS-9-protein precu
4	140.5	8.6	527	2	A32469	80K protein H prec
5	106	6.5	150	2	T01158	hypothetical prote
6	95	5.8	407	2	F70318	hypothetical prote
7	93.5	5.7	216	2	A48343	24K hypothetical p
8	93.5	5.7	136	2	F86904	conserved hypotet
9	92	5.6	515	2	T43152	hypothetical prote
10	92	5.6	544	1	I36911	involucrin L - dou
11	92	5.6	1194	2	CS9436	KIRAI191 protein l
12	91.5	5.6	216	2	A39926	hypothetical prote
13	91.5	5.6	506	2	T41623	probable protein k
14	91.5	5.6	1014	2	T30545	major surface glyco
15	91	5.6	2049	2	T47587	TATA box binding p
16	90.5	5.5	635	1	I37060	involucrin L - gor
17	90	5.5	495	1	I37062	involucrin S - gor
18	90	5.5	5262	2	T03454	ALR protein - huma
19	89	5.4	2025	2	JC5020	tetratricopeptide
20	88	5.4	457	2	S66270	kynurenine-oxogluc
21	88	5.4	935	2	T51930	kinesin [imported]
22	88	5.4	1328	2	SS2481	tyb protein - yeas
23	87.5	5.4	364	2	B83078	probable D-amino a
24	87.5	5.4	605	1	I37061	involucrin M - gor
25	87.5	5.4	26926	1	I38344	titin, cardiac mus
26	87	5.3	638	2	B84799	similar to axi 1 p
27	87	5.3	891	2	T30812	ubiquitin-protein
28	86.5	5.3	493	1	A57783	involucrin - cotto
29	86	5.3	816	2	B86285	hypothetical prote

30	86	5.3	934	2	T25864	hypothetical prote
31	85.5	5.2	320	2	H95233	L-asparaginase, pr
32	85.5	5.2	320	2	B98098	asparaginase (Ec 3
33	85.5	5.2	560	1	A40125	involucrin - pygmy
34	85.5	5.2	713	2	UC2522	nuclear autoantige
35	85.5	5.2	2128	2	I52577	beta-spectrin - mo
36	85	5.2	704	1	S33704	transforming prote
37	85	5.2	1192	2	T16611	probable serine/th
38	85	5.2	1246	2	G89287	protein H39E23.1 [
39	84.5	5.2	364	2	T23819	hypothetical prote
40	84.5	5.2	678	2	C72819	flagellar biosynth
41	84.5	5.2	1174	2	T50973	probable alpha-aml
42	84.5	5.2	4377	2	A55875	ankyrin 3, long sp
43	84	5.1	813	2	AC3304	non-motile and pha
44	84	5.1	2677	2	A38194	desmoplakin I - hu
45	83	5.1	354	2	T04779	hypothetical prote

ALIGNMENTS

RESULT 1
T45062
hypothetical protein c316G12.3 [imported] - human
C/Species: Homo sapiens (man)
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #ext_change 21-Jan-2000
C/Accession: T45062
R/Frankland, J.
submitted to the EMBL Data Library, July 1999
A/Reference number: Z22901
A/Accession: T45062
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-305 <FRA>
A/Cross-references: EMBL:AL031709; PIDN:CA56184.1
A/Experimental source: clone LA16-316G12
C/Genetics:
A/Intons: 18/1; 37/2; 60/1; 78/2; 106/2; 137/3; 176/1; 203/3; 247/3; 275/1
A/Note: c316G12.3

Query Match 100.0%; Score 1634; DB 2; Length 305;
Best Local Similarity 100.0%; Pred. No. 3e-132;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAAGLARLLILGLISAGGPA	MAAGKAYEENPNA	GVNNP	PQASRIQAKR	DPSPVG	60
DB	1	MAAGLARLLILGLISAGGPA	MAAGKAYEENPNA	GVNNP	FLPQASRIQAKR	DPSPVG	60
QY	61	PVHLFRLSGKCFSLVSTY	KYECPPHNTY	QHROTFRMNA	YSGLIGIMHWEI	ANNTFTG	120
DB	61	PVHLFRLSGKCFSLVSTY	KYECPPHNTY	QHROTFRMNA	YSGLIGIMHWEI	ANNTFTG	120
QY	121	MMNRDGDGCRSRSRQSK	VELACGKSNRLAHV	SPSTCYVALT	ETPLVCHPHALV	YPTL	180
DB	121	MMNRDGDGCRSRSRQSK	VELACGKSNRLAHV	SPSTCYVALT	ETPLVCHPHALV	YPTL	180
QY	181	PEALQRMQDVEQDLDEL	ITPQGHKRLRTLE	PDAGYLT	KPENEP	TOLEGSPDSIGFE	240
DB	181	PEALQRMQDVEQDLDEL	ITPQGHKRLRTLE	PDAGYLT	KPENEP	TOLEGSPDSIGFE	240
QY	241	TLENCRANKELSKEIKR	LKGLLTQHGIPYTR	PTETSNTEHLGHET	PRAKSPQRLGDPG	300	
DB	241	TLENCRANKELSKEIKR	LKGLLTQHGIPYTR	PTETSNTEHLGHET	PRAKSPQRLGDPG	300	
QY	301	LRGSL	305				
DB	301	LRGSL	305				

RESULT 2
T24944
hypothetical protein ZK1307.8 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 23-Oct-1999
C/Accession: T24944, T27735
R/Gardner, A.
Submitted to the EMBL Data Library, January 1995
A/Reference number: Z19958
A/Accession: T24944
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-507 <M12>
A/Cross-references: EMBL:Z47356, PIDD:CAA87420.1, GSPDB:GN00020, CESP:ZK1307.8
A/Experimental source: clone T15H9
R/Berks, M.
Submitted to the EMBL Data Library, January 1995
A/Accession: T27735
A/Reference number: Z20412
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-507 <M12>
A/Cross-references: EMBL:Z47358, PIDD:CAA87438.1, GSPDB:GN00020, CESP:ZK1307.8
C/Genetics:
A/Map position: 2
A/Map position: 34/1, 177/1, 265/1, 463/3
Query Match 11.4%; Score 185.5; DB 2; Length 507;
Best Local Similarity 38.7%; Pred. No. 6.7e-08;
Matches 41; Conservative 11; Mismatches 47; Indels 7; Gaps 4;
QY 67 LSGKPSLVESTYKTEFCFPHNVTOEOTFRNNAVSGILGIWHEWE-IAANTFTGMNR 124
DB 394 LDKCFEDRVQYTYQFCFGQNTQKT---GAYSGTSLSPFKWMSGPEGNKYSKHFG 449
QY 125 DSDAC-RSRRSQKVELACGKSNRLAHVSEPTCYVALTFETPLVC 169
DB 450 DQGCQMGKPKRSTDTIBGEMNELVEYTERPAKCYLTFRTPLAC 495
RESULT 3
JC5889
OS-9 protein precursor - human
N/contains: OS-9 protein, splice form 1; OS-9 protein, splice form 2; OS-9 protein, splice form 3
C/Species: Homo sapiens (man)
C/Date: 18-Mar-1998 #sequence_revision 18-Mar-1998 #text_change 22-Oct-1999
R/Accession: JC5889, JE0106, JE0107, JE0108
R/Kimura, Y.; Nakazawa, M.; Tsuchiya, N.; Asakawa, S.; Shimizu, N.; Yamada, M.
J. Biochem. 122, 1190-1195, 1997
A/Title: Genomic organization of the OS-9 gene amplified in human sarcomas.
A/Reference number: JC5889; MUID:98158329; PMID:9498564
A/Accession: JC5889
A/Molecule type: DNA
A/Residues: 1-667 <KIM1>
A/Cross-references: DDBJ:AB002806, NID:G2780782, PIDD:BA24363.1, PID:dl025275; PID:G2780782
A/Experimental source: sarcomas
A/Note: neither the complete nucleic acid sequence nor the complete translation are shown
R/Kimura, Y.; Nakazawa, M.; Yamada, M.
J. Biochem. 123, 876-882, 1998
A/Title: Cloning and characterization of three isoforms of OS-9 cDNA and expression of cDNA
A/Reference number: JE0106, MUID:98230694; PMID:9562620
A/Accession: JE0106
A/Molecule type: mRNA
A/Residues: 1-667 <KIM2>
A/Cross-references: DDBJ:AB002806, NID:G2780782, PIDD:BA24363.1, PID:dl025275; PID:G2780782
A/Accession: JE0107
A/Molecule type: mRNA
A/Residues: 1-534, 590-667 <KIM3>
A/Cross-references: DDBJ:AB002806
A/Accession: JE0108
A/Molecule type: mRNA
A/Residues: 1-455, 471-534, 590-667 <KIM4>
A/Cross-references: DDBJ:AB002806
C/Comment: This protein is involved in amplification and overexpression of various tumor C/Genetics:

A/Genes: OS-9
A/Cross-references: GDB:9958646
A/Map position: 12q13-12q15
A/Keywords: alternative splicing; carcinogenesis; glycoprotein
F/1-25/Domain: signal sequence #status predicted <Sig>
F/26-667/Product: OS-9 protein, splice form 1 #status predicted <MAT1>
F/26-534, 590-667/Product: OS-9 protein, splice form 2 #status predicted <MAT2>
F/26-455, 471-534, 590-667/Product: OS-9 protein, splice form 3 #status predicted <MAT3>
F/177/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 9.5%; Score 155.5; DB 2; Length 667;
Best Local Similarity 21.7%; Pred. No. 3.6e-05;
Matches 79; Conservative 47; Mismatches 121; Indels 117; Gaps 18;
QY 21 PAGAAMKVEEENACVNNPFLPQASRLQARDBSPVGPVHLRLSKCSLYESTYK 80
DB 76 PAGAHPQEREESTPAYQGPGLPEL-LSPMR-----APCLTKTDWWT 119
QY 81 YEECFPHNVTO-----HEOTFRNNAVSGILGIWHEWEIAANTFTGMNR 123
DB 120 YEECYGRHITQYMEDESEIKGEVLYGYQSAFMDDETERAKSKGHRKRYHSQTYG--- 176
QY 124 RQGDAC-RSRRSQKVELACGK-----SNRLAHVSEPTCYVALTFETPLVC----- 169
DB 177 -NSKCDLNGRPREAEVREPLCDEGAGISGDYIDRVDPLSCSYVLTIRPRLCPPLLRP 235
QY 170 ---HPHALLVPTL-PEA-----LORQWDVE-----QDLDEL-----ITPO- 203
DB 236 PPSAALQALICHPDSLOPEEYMAVQGRADSKQYGDKIIELODLGPQWSEKSGVAQK 295
QY 204 -----GHEKLRTLFEDA-GYLTPE-ENEPT-----QLEGPSLGFTL 242
DB 296 MAGASPTKDSKSDPFWKMLNPEEDQAPGGEVPEABDPPSEADSAAGAPNDQRNVQ 355
QY 243 ENCRKAHKLSKIKELKGLTQHGIPYRPRETSTNLTGHEPT-----RAASPOL 295
DB 356 VKYIRSPADLIRIIEELKG-CTKKGK-----NIGQGVYDDAAEVPQREPK 403
QY 296 RQDP 299
DB 404 RQDP 407
RESULT 4
A32469
80K protein H precursor - human
C/Species: Homo sapiens (man)
C/Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 21-Jul-2000
R/Accession: A32469
R/Sakai, K.; Hirai, M.; Minoshima, S.; Kudoh, J.; Fukuyama, R.; Shimizu, N.
Genomics 5, 309-315, 1989
A/Title: Isolation of cDNAs encoding a substrate for protein kinase C: nucleotide sequence
A/Reference number: A32469; MUID:90007553; PMID:2793184
A/Accession: A32469
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-527 <SAK>
A/Cross-references: GB:J03075, NID:G182854, PIDD:AAA52493.1, PID:G182855
A/Accession: A32469
A/Genetics:
A/Genes: GDB:PRKCSH, G19P1
A/Cross-references: GDB:119961; OMIM:177060
A/Map position: 19p13.2-19p13.3
C/Keywords: phosphoprotein
F/1-14/Domain: signal sequence #status predicted <Sig>
F/524-527/Region: endoplasmic reticulum retention signal
Query Match 8.6%; Score 140.5; DB 2; Length 527;
Best Local Similarity 30.5%; Pred. No. 0.00051;
Matches 32; Conservative 14; Mismatches 54; Indels 5; Gaps 3;
QY 67 LSGKPSLVESTYKTEFCFPHNVTOEOTFRNNAVSGILGIWHEWE-IAANTFTGMNR 125
DB 410 LYSQCELTNTNEYVRLCPFKLVSQP---XUGSPSTLGTGWSIGPDHDKFSAMKYEQ 466

A/Residues: 1-1436 <STO>
 A/Cross-references: GB:AF005176; PID:g12725306; PIDN:AAK06336.1; GSPDB:GN00146
 A/Experimental source: strain IL403
 C/Genetics:
 A/Gene: yxcA

Query Match 5.7%; Score 93.5; DB 2; Length 1436;
 Best Local Similarity 27.3%; Pred. No. 21;
 Matches 47; Conservative 21; Mismatches 55; Indels 49; Gaps 9;
 QY 115 NNTFTGMWRDGD--ACSRSRQSKVELACGSKNRLAHVSEPTCVYALTFETPL-----167
 DB 959 NCLIVYVKIDEGSNMGAVIRLRSLKAAV---SERVRNITLASTVEHLYQETPAFTKE 1014
 QY 168 VCHPHALVYPTLPALQKQMDVODLADDELITPQGHKRLTLFEDAGY--LKTPEEN 225
 DB 1015 MAKKATLL---LP-----MLSTIHOEGILDTFAAAGNIVSLPSBN 1053
 QY 226 EFTQLEGPDLSGFETLNCRAKHELSKEIKELG-----LITQHG 267
 DB 1054 --TSVNNG--LKFVNNDSCTPAITITIGQLIALQSGEYDLDNSTVMMTQTG 1100

RESULT 9
 T43152
 hypothetical protein - fission yeast (Schizosaccharomyces pombe) (fragment)
 C/Species: Schizosaccharomyces pombe
 C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C/Accession: T43152
 R.Yoshioaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
 DNA Res. 4, 363-369, 1997
 A/Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
 A/Reference number: Z17323; MUID:98162722; PMID:9501991
 A/Accession: T43152
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-515 <YOS>
 A/Cross-references: EMBL:D89245; NID:g1749697; PIDN:BAAL3906.1; PID:g1749698
 A/Experimental source: strain PR745

Query Match 5.6%; Score 92; DB 2; Length 515;
 Best Local Similarity 21.9%; Pred. No. 7;
 Matches 25; Conservative 16; Mismatches 47; Indels 26; Gaps 3;
 QY 87 HNVTHQEGTFR--NNAYSGLIGIWHWEIANNTEFLMG-----122
 DB 373 HTKMLHELYHMGWLYRAIKMETIKELIGTYVYEVVPODSILGNFASQEGNVL 432
 QY 123 -MRDGDAC-RSRSRQSKVELACGSKNRLAHVSEPTCVYALTFETPLVCHPHAL 174
 DB 433 KYENGQSCWNGPHRSALIVTEGCVENEIVSVLEAOKCEYLKMKSPAACSPNQL 486

RESULT 10
 136911
 involucrin L - douroucoul
 C/Species: Aotus trivirgatus (douroucoul; night monkey, owl monkey)
 C/Date: 16-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 22-Jun-1999
 C/Accession: 136911
 R.Tsang, H.; Green, H.
 Mol. Biol. Evol. 6, 460-468, 1989
 A/Title: The involucrin gene of the owl monkey: origin of the early region.
 A/Reference number: 136911; MUID:90014142; PMID:2507864
 A/Accession: 136911
 A/Status: nucleic acid sequence not shown
 A/Molecule type: DNA
 A/Residues: 1-544 <RES>
 A/Cross-references: GB:M25313; NID:g176557; PIDN:AAA35375.1; PID:g176558
 C/Comment: During the terminal differentiation of keratinocytes, this protein from the c
 linked envelope under the plasma membrane.
 C/Species: involucrin
 C/Keywords: cornified cell envelope; duplication; epidermis; tandem repeat
 P.153-501/Region: 10-residue repeats (Q-E-G-Q-[PLV]-[KH]-[PL]-E-Q)

Query Match 5.6%; Score 92; DB 1; Length 544;
 Best Local Similarity 23.4%; Pred. No. 7.5;
 Matches 34; Conservative 25; Mismatches 48; Indels 38; Gaps 4;
 QY 182 EALQKQMDVODLADL-----LITPQGHKRLTLTFED 215
 DB 110 EKQKREKQQLQGLBEKELLDQPDHLEAKSDEQLGTKEQLLEPPEQOEQLKCLEGO 169
 QY 216 AGYLTKEENEPITQLEGPDLSGFETLNCRAKHELSKEIK-RKGLITGIGIPYTRPT 274
 DB 170 EGHLEPQOEBO-----LKCLEO-QEGHLEPQOEBOQLNHEQOQLKHEBO 218
 QY 275 ETSNLEHLEHTPRAPKSPQRLKDP 299
 DB 219 QEGVAKHLEQOEKSELPEQORQGP 243

RESULT 11
 C59436
 KIAA1391 protein [imported] - human
 C/Species: Homo sapiens (man)
 C/Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 23-Sep-2002
 C/Accession: C59436; D59436
 R.Nagase, T.; Kikuno, R.; Ishikawa, K.I.; Hirose, M.; Ohara, O.
 DNA Res. 7, 65-73, 2000
 A/Title: Prediction of the coding sequences of unidentified human genes. XVI. The compl
 A/Reference number: C59436
 A/Accession: C59436
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1194 <NAG>
 A/Cross-references: GB:BA92629; PID:g7243163; PIDN:BA92629.1
 R. Ohara, O.; Nagase, T.; Kikuno, R.
 submitted to Genbank, January 2000
 A/Reference number: D59436
 A/Accession: D59436
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1194 <OHA>
 A/Cross-references: GB:BA92629; PID:g7243163; PIDN:BA92629.1

Query Match 5.6%; Score 92; DB 2; Length 1194;
 Best Local Similarity 22.2%; Pred. No. 22;
 Matches 63; Conservative 40; Mismatches 103; Indels 78; Gaps 13;
 QY 14 LSNAGPAPAGA-----AKYKVEEPNMAFGVNNPFLPQASRLQAR---DSEP-VSGPVH 63
 DB 941 LSPGTPSPGSSVSSQDSAFSQISEHSYFTPTETSSPIDCTFOAKRKEDLSPDSNASH 1000
 QY 64 LFRLL---SGKCFSLVESYKYEF---CFHNVTQHEOTFRNMAVSGILGIMHEIANN 116
 DB 1001 VSGMPEPSSGQACSRPATYTKDVTMWHSGMHSVTLHPST-----1039
 QY 117 FTFGMMWRDGD--ACSRSRQSKVELACGSKNRLAHVSEPTCVYALTFETPLVCHPHAL 175
 DB 1040 -----WLRNGVSLNKNWSLKKAKAARPEEEKIAPKXP-----LEPP--PHA-- 1080
 QY 176 VFTPLPEALQKQMDVODL--ADELITPQGHKRLTLFEDAGYLTKEENEPITQLEG 233
 DB 1081 --SGVPEANSLQEBQDPLPRAEGLSPVQASQSSSPFDOS-----ERHCSS 1127
 QY 234 PSLGFETLNCRAKHELSKEIKRLKGLITGIGIPYTRPTTS 277
 DB 1128 PSLVESRLKLCMKSHHEIPESQSSG-----SLPWRASASS 1166

RESULT 12
 A33926
 hypothetical protein 1 - chicken anemia virus
 C/Species: chicken anemia virus, CAV
 C/Date: 17-Jan-1992 #sequence_revision 17-Jan-1992 #text_change 08-Oct-1999
 C/Accession: A33926


```

Db      976 APTKDIPLPYAELSTYTKM-RNEASQILHIVE-TCHCFDKLSTNKLN---VESVTAD 1030
Qy      93  EQTFRNNAYSGLIGIWEHEIANTFTGMMWBDGACRSRSQSKVELACGKNRLAHVS 152
Db      1031 ETT---DPASTLDLNNKESAGNESLEKOVFEDEVESRQQLISTAGYLKCVQSN--LHIT 1084
Qy      153 EPSTCVYALTF--ETPLVCHPHALIVPTLPALOROWDOVEODLADLIT----- 201
Db      1085 VTSLVAAAVVWMSBPAPLNP---ITLP-LMASIKREOQLQIJAELAEALAYCVDR 1140
Qy      202 -PQHEKLI RT-----LFEDAGYL---KTPENEPYOLEGGP 234
Db      1141 KPSPNDKLIKNI CSLTCMDPSETPOASIISSMDIVDDMDFLSSRSNTGKQAKAVVLASGE 1200
Qy      235 DSL---GPELLENCRAHKELS-----KEIKRLKGLITQHGIP 269
Db      1201 DRSKYEGFTTRGSEALKHLSLKFGGSLFDKLPKLMECTIEVLVP 1246

```

Search completed: July 26, 2004, 11:12:26
 Job time : 9.22934 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:58 ; Search time 5.08008 Seconds
(without alignments)
3126.212 Million cell updates/sec

Title: US-10-657-280-3
Perfect score: 1634
Sequence: 1 MAAGIARLLLLIGSAGPA.....TPRAKSPQRLGRDGLGSL 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	155.5	9.5	667	1 OS9_HUMAN	Q1438 homo sapien
2	141	8.6	533	1 G19P_BOVIN	Q28034 bos taurus
3	140.5	8.6	527	1 G19P_HUMAN	P14314 homo sapien
4	115.5	7.1	377	1 K2L4_HUMAN	Q99706 homo sapien
5	95.5	5.8	216	1 VP2_CAVC1	Q99151 chicken ane
6	94.5	5.8	216	1 VP2_CAV82	P54093 chicken ane
7	92.5	5.7	216	1 VP2_CAV26	P54092 chicken ane
8	92	5.6	544	1 INV0_AOTTR	P24708 actus trivi
9	90	5.5	5262	1 ML22_HUMAN	O14686 homo sapien
10	89	5.4	2025	1 TTC3_HUMAN	P58804 homo sapien
11	88	5.4	593	1 STB2_CANFA	Q28288 canis faml1
12	88	5.4	935	1 KINH_SYNRA	O43093 syncephalas
13	88	5.4	1328	1 YMD9_YEAST	Q03434 saccharomyc
14	87.5	5.4	364	1 YPE8_PSEAE	P33642 pseudomonas
15	87.5	5.4	703	1 SYM_CHLTE	P59077 chlorobium
16	87	5.3	445	1 PLI3_MOUSE	Q89X76 mus musculu
17	87	5.3	469	1 PLI3_HUMAN	Q842b7 mus musculu
18	86.5	5.3	246	1 DEOC_THEAC	Q942b7 homo sapien
19	86.5	5.3	493	1 INVO_SAGOE	P24712 saguinus oe
20	85.5	5.2	560	1 INVO_PANPA	P14591 pan paniscu
21	85.5	5.2	797	1 STR3_HUMAN	Q10033 homo sapien
22	85.5	5.2	2128	1 SPGB_MOUSE	P15508 mus musculu
23	85	5.2	704	1 MYRB_MOUSE	P48972 mus musculu
24	84.5	5.2	4377	1 ANK3_HUMAN	Q13953 homo sapien
25	84	5.1	2871	1 DESP_HUMAN	P15924 homo sapien
26	83	5.1	1248	1 SM1B_MOUSE	Q92046 mus musculu
27	82.5	5.0	286	1 PUJ1_SCICO	P22311 sciara copr
28	82.5	5.0	906	1 HELY_MYCTU	Q10701 mycobacteri
29	81.5	5.0	322	1 PGJB_HUMAN	Q09646 homo sapien
30	81.5	5.0	383	1 PYRC_PYRAE	Q842c7 pyrobaculum
31	81.5	5.0	406	1 ZYJ3_HUMAN	Q96167 homo sapien
32	81.5	5.0	819	1 M4K2_HUMAN	Q1851 h mitogen-a
33	81	5.0	322	1 PGJB_MOUSE	P70186 mus musculu

34	81	5.0	579	1 MDIA_BUCBP	O89a97 buchnera ap
35	81	5.0	897	1 TBEC_HUMAN	O95759 homo sapien
36	81	5.0	1004	1 PTPX_RAT	Q63475 rattus norv
37	80.5	4.9	473	1 HYA2_RAT	Q924q3 rattus norv
38	80.5	4.9	835	1 INVO_PONPY	P14708 pongo pygma
39	80	4.9	279	1 ATPG_MYCPN	Q50330 mycoplasma
40	80	4.9	321	1 PGJB_BOVIN	P79119 bos taurus
41	80	4.9	410	1 VEE2_BPT1	P03122 bovine papl
42	80	4.9	770	1 GLOB_SYNY3	P52881 synchocyst
43	80	4.9	1328	1 YME4_YEAST	Q04711 saccharomyc
44	80	4.9	1328	1 YMT5_YEAST	Q04214 saccharomyc
45	80	4.9	1328	1 YMU0_YEAST	Q04670 saccharomyc

ALIGNMENTS

RESULT 1
ID OS9_HUMAN STANDARD; PRT; 667 AA.
AC Q1438; 000579;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Protein OS-9 precursor.
GN OS9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96198566; PubMed=8634085;
RA Su Y.A., Hutter C.M., Trent J.M., Meltzer P.S.;
RT "Complete sequence analysis of a gene (OS-9) ubiquitously expressed
RT in human tissues and amplified in sarcomas.";
RL Wcl. Carcinog. 15:270-275 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98158329; PubMed=9498564;
RA Kimura Y., Nakazawa M., Tsuchiya N., Asakawa S., Shimizu N.,
RA Yamada M.;
RT "Genomic organization of the OS-9 gene amplified in human sarcomas.";
RL J. Biochem. 122:1190-1195 (1997).
RN [3]
RP SEQUENCE FROM N.A.; CHARACTERIZATION, AND ALTERNATIVE PRODUCTS.
RC TISSUE=Promyelocytic leukemia;
RX MEDLINE=98230694; PubMed=9562620;
RA Kimura Y., Nakazawa M., Yamada M.,
RT "Cloning and characterization of three isoforms of OS-9 cDNA and
RT expression of the OS-9 gene in various human tumor cell lines.";
RL J. Biochem. 123:876-882 (1998).
RN [4]
RP SEQUENCE OF 194-667 FROM N.A.
RX MEDLINE=97336055; PubMed=9192850;
RA Elkahoun A.G., Krizman D.B., Wang Z., Hofmann T.A., Roe B.,
RA Meltzer P.S.;
RT "Transcript mapping in a 46-kb sequenced region at the core of 12q13.3
RT amplification in human cancers.";
RL Genomics 42:295-301 (1997).
CC -!- FUNCTION: May play a role in cell viability.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=OS-9-1;
CC IsoId=Q1438-1; Sequence=Displayed;
CC Name=OS-9-2;
CC IsoId=Q1438-2; Sequence=VSP_004353;
CC Name=OS-9-3;
CC IsoId=Q1438-3; Sequence=VSP_004352, VSP_004353;
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed. Found as well in all
CC tumor cell lines analyzed, amplified in sarcomas. Highly expressed
CC in osteosarcoma OSA-CL and rhabdomyosarcoma RH30 cell lines. OS-9-
CC 2 is the major isoform detected in all cell types examined.

RL Genomics 5:309-315 (1989).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97196775; PubMed=9043864;
 RA Ophoff R.A., Terwindt G.M., Verschuwe M.N., van Eijk R.,
 RA Mohrenweiser H., Litt M., Hofker M.H., Haan J., Ferrari M.D.,
 RA Francis R.R.;
 RT "A 3-Mb region for the familial hemiplegic migraine locus on 19p13.1-
 RT p13.2: exclusion of PRKCSH as a candidate gene.";
 RL Eur. J. Hum. Genet. 4:321-328 (1996).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J03075; AAA52493.1; -
 DR EMBL: U50326; AAA98668.1; -
 DR EMBL: U50317; AAA98668.1; JOINED.
 DR EMBL: U50318; AAA98668.1; JOINED.
 DR EMBL: U50319; AAA98668.1; JOINED.
 DR EMBL: U50320; AAA98668.1; JOINED.
 DR EMBL: U50321; AAA98668.1; JOINED.
 DR EMBL: U50322; AAA98668.1; JOINED.
 DR EMBL: U50323; AAA98668.1; JOINED.
 DR EMBL: U50324; AAA98668.1; JOINED.
 DR EMBL: U50325; AAA98668.1; JOINED.
 DR PIR: A32469; A32469.1; JOINED.
 DR Genem: HGNC:9411; PRKCSH.
 DR MIM: 177060; -
 DR GO: GO:0005622; C:intracellular; NAS.
 DR GO: GO:0007243; P:protein kinase cascade; NAS.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000886; ER_target_5.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR009011; Man_6_P_R_bird.
 DR Pfam: PF00036; efhand; 2.
 DR SMART: SM00192; LDLr; 1.
 DR PROSITE: PS00014; ER_TARGET; 1.
 DR PROSITE: PS00018; EF_HAND; 1.
 DR KX Phosphorylation.
 FT DOMAIN 313 336 GLU-RICH (ACIDIC).
 FT MOD_RES 89 89 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 FT MOD_RES 382 382 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 FT MOD_RES 389 389 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 FT MOD_RES 433 433 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 SC SEQUENCE 527 AA; 59296 MW; 2BB2F9AC14E3FC94 CRC64;
 Query Match 8.6%; Score 140.5; DB 1; Length 527;
 Best Local Similarity 30.5%; Pred. No. 0.00033;
 Matches 32; Conservative 14; Mismatches 54; Indels 5; Gaps 3;
 QY 67 LSGKCFSLVSTYKKEPCFPHNVHTQETFRNNAYSGILGIWHEN-BIANTFTGMWRD 125
 DB 410 LYSQCYELTNYEYVRLCPFKLVSKP---KLGGSPSLGTWGSWIGPDHDKFSAMKXEQ 466
 QY 126 GDAC-RSRSRQSKVELLACGSKNRLAHVSEPSVCYALTFEPPLVC 169
 DB 467 GTCGWQGPFRSTVRLTLCGKETMTSTSPSKCEYLMELTPAAC 511
 ID K214 HUMAN STANDARD; PRT: 377 AA.
 AC Q99706; Q14622; Q14623; Q14624; Q43534; P78400; P78401;
 AC Q99559; Q99560; Q99561; Q99562; Q90077;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 10-OCT-2003 (Rel. 42; Last annotation update)
 DE Killer cell immunoglobulin-like receptor 2D14 precursor (MHC class I

DE NK cell receptor KIR103AS) (Killer cell inhibitory receptor 103AS)
 DE (KIR-103AS) (G9P).
 GN KIR2DL4 OR KIR103AS.
 OS Homo sapiens (human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 CC NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97102173; PubMed=8946682;
 RA Selvakumar A., Steffens U., Dupont B.;
 RT "NK cell receptor gene of the KIR family with two IG domains but
 RT highest homology to KIR receptors with three IG domains";
 RL Tissue Antigens 48:285-294 (1996).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 3; 4 AND 5).
 RX MEDLINE=97378880; PubMed=9234477;
 RA Selvakumar A., Steffens U., Palanisamy N., Chaganti R.S.K., Dupont B.;
 RT "Genomic organization and allelic polymorphism of the human killer
 RT cell inhibitory receptor gene KIR103.";
 RL Tissue Antigens 49:564-573 (1997).
 RN (3)
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 6).
 RC TISSUE=Lymphoid;
 RA Biasoni R.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN (4)
 RP SEQUENCE FROM N.A. AND VARIANTS.
 RX MEDLINE=98090086; PubMed=9430221;
 RA Ueberberg M., Valliant N.M., Shum B.P., Shilling H.G.,
 RA Llenert-Weidenbach K., Corliss B., Tyan D., Lanier L.L., Parham P.;
 RT "Human diversity in killer cell inhibitory receptor genes.";
 RL Immunity 7:753-763 (1997).
 RN (5)
 RP SEQUENCE FROM N.A.
 RA Chan H.W., Salter R.D.;
 RT "Exon deletion contributes to structural diversity of 2D14 killer
 RT inhibitory receptors.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN (6)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=97213129; PubMed=9059894;
 RA Selvakumar A.;
 RL "Polymorphism and domain variability of human killer cell inhibitory
 RL receptor.";
 RT Immunol. Rev. 155:183-196 (1997).
 CC -1- FUNCTION: RECEPTOR ON NATURAL KILLER (NK) CELLS FOR HLA-C ALLELES.
 CC INHIBITS THE ACTIVITY OF NK CELLS THUS PREVENTING CELL LYSIS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=6;
 CC Name=1;
 CC IsoId=Q99706-1; Sequence=displayed;
 CC Name=2; Synchronms=AST;
 CC IsoId=Q99706-2; Sequence=VSP_002609;
 CC Name=3; Synchronms=AS;
 CC IsoId=Q99706-3; Sequence=VSP_002610;
 CC Name=4; Synchronms=ASDI;
 CC IsoId=Q99706-4; Sequence=VSP_002609, VSP_002610;
 CC Name=5; Synchronms=ASD2;
 CC IsoId=Q99706-5; Sequence=VSP_002609, VSP_002610, VSP_002611;
 CC Name=6;
 CC IsoId=Q99706-6; Sequence=VSP_002608, VSP_002609, VSP_002610;
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL, U711199; AAB49756.1; -.
DR EMBL, AF003123; AAB61926.1; -.
DR EMBL, AF003117; AAB61926.1; JOINED.
DR EMBL, AF003117; AAB61926.1; JOINED.
DR EMBL, AF003118; AAB61926.1; JOINED.
DR EMBL, AF003119; AAB61926.1; JOINED.
DR EMBL, AF003121; AAB61926.1; JOINED.
DR EMBL, AF003122; AAB61926.1; JOINED.
DR EMBL, AF003120; AAB61926.1; JOINED.
DR EMBL, AF002979; AAB71387.1; -.
DR EMBL, AF002980; AAB71388.1; -.
DR EMBL, AF002981; AAB71389.1; -.
DR EMBL, AF002982; AAB71390.1; -.
DR EMBL, X99479; CAA67842.1; -.
DR EMBL, X99480; CAA67843.1; -.
DR EMBL, AF034771; AAB95164.1; -.
DR EMBL, AF034772; AAB95165.1; -.
DR EMBL, AF034773; AAB95166.1; -.
DR EMBL, AF110033; AAD24763.1; -.
DR EMBL, AF110032; AAD24763.1; JOINED.
DR EMBL, AF110033; AAD24763.1; JOINED.
DR EMBL, AF110034; AAD24763.1; JOINED.
DR EMBL, U73394; AAC51146.1; -.
DR Gene, HGNC:6332; KIR2DL4.
DR WIM, 604945; -.
DR GO, GO:0005887; C:integral to plasma membrane; TAS.
DR GO, GO:0004888; F:transmembrane receptor activity; TAS.
DR GO, GO:0006968; P:cellular defense response; TAS.
DR GO, GO:0007165; P:signal transduction; TAS.
DR InterPro, IPR007110; Ig-like.
DR Pfam, PF00047; Ig_2.
DR PROSITE, PS00835; IG-LIKE, FALSE NEG.
KW Receptor; Immunoglobulin domain; Glycoprotein; Signal; Transmembrane;
  Repeat; Multisene family; Alternative splicing; Polymorphism.
FT SIGNAL 1 21
FT CHAIN 22 377
FT DOMAIN 22 242
FT TRANSMEM 243 263
FT DOMAIN 264 377
FT DOMAIN 44 104
FT DOMAIN 139 202
FT DISULFID 51 97
FT CARBOHYD 146 195
FT CARBOHYD 141 141
FT CARBOHYD 175 175
FT VARSPLIC 27 121
FT VARSPLIC 219 235
FT FT 235
FT FT 270
FT VARSPLIC 236 270
FT FT 270
FT VARSPLIC 271 288
FT VARSPLIC 53 53
FT VARSPLIC 87 87
FT VARSPLIC 138 138
FT VARSPLIC 209 209
FT VARSPLIC 271 271
FT VARSPLIC 371 371
FT CONFLICT 229 229
FT SEQUENCE 377 AA; 41527 MW; 7ECADF0FC7F4143B CRC64;

```

```

Query Match 7.1%; Score 115.5; DB 1; Length 377;
Best Local Similarity 21.3%; Pred. No. 0.027;
Matches 69; Conservative 43; Mismatches 117; Indels 95; Gaps 15;

QY 19 PAPAAGAAKKV--EEPNAFGV--NNPFLQARLQAKDPSVSGPV----- 62
DB 89 PAAAGTYRRCRGPHPSPTWMSAPNDLVMTVGLVKEBGLTARPPPTVRTGENVTLLSCSS 148
QY 63 -----HFLRSGKCFSL-----VESTYKFECPFNNTVQHQTFRMNAVSGILGIW 108
DB 149 QSSFDLYHLSR--EGEHEHLRLPAVPSINQTFQADP--PLGPAI--HGETYR--CFGSPHSGP 203
QY 109 HEWEIANNFTFGKWMEDGACRSRQSVELACGSKNRLAVSEPSVTVLTFETPLV 168
DB 204 YEWSDASDPL-----FVSYTGNNPSSSWPSPPTBPS-----FKTGIA 238
QY 169 CHPH-----ALLVPTLPEALQRCQWQDVEDLADDELITPGHEKLT--RTLPEDAGYKLT 221
DB 239 RHILHAVIRYSVALIIFTLIPFELIRWCKSKDDAVMNPAGHRTVNR---EDSD----- 291
QY 222 PEENEPTOLEGGPDSIGFETLNCRAHKEISKIRLKGILTOHGIPYTRP--TETSNI 280
DB 292 -----EOPQOEVTYAQLDHCIFTOQRKINGPSQRSK-----RSTDTSVCI 331
QY 281 HIGHEPRAPKSPPEQLRGDPRLGS 304
DB 332 ELPNAEPRLSPAHHEHSQLMGSS 355

RESULT 5
ID VP2_CAVCI STANDARD; PRT; 216 AA.
AC Q99151; P54091;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Protein VP2.
OS Chicken anemia virus (German isolate Cuxhaven-1) (CAV), and
OC Chicken anemia virus (USA isolate CIA-1) (CAV).
OC Viruses; ssDNA viruses; Circoviridae; Gyrovirus.
OX NCBI_TaxID=73475; 73478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Cuxhaven-1;
RX MEDLINE=91237831; PubMed=1851873;
RA Noteborn M.H.M., de Boer G.F., van Roozelaar D.J., Karreman C.,
  Kraendburg O., Vos J.G., Jeurissen S.H.M., Hoebe R.C., Zanenema A.,
  Koch G., van Ormondt H., van der Bo A.J.;
RT "Characterization of cloned chicken anemia virus DNA that contains
  all elements for the infectious replication cycle.";
RL J. Virol. 65:3131-3139(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Cuxhaven-1;
RX MEDLINE=9226898; PubMed=1605740;
RA Meenan B.M., Todd D., Creelan J.L., Earle J.A.P., Hoey E.M.,
  McNulty M.S.;
RT "Characterization of viral DNAs from cells infected with chicken
  anaemia agent: sequence analysis of the cloned replicative form and
  transfection capabilities of cloned genome fragments.";
RL Arch. Virol. 124:301-319(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CIA-1;
RX MEDLINE=97126092; PubMed=8971016;
RA Renshaw R.W., Soine C., Weinkle T., O'Connell P.H., Ohashi K.,
  Watson S., Lucio B., Harrington S., Schat K.A.;
RT "A hypervariable region in VP1 of chicken infectious anemia virus
  mediates rate of spread and cell tropism in tissue culture.";
RL J. Virol. 70:8872-8878(1996).
RN [4]
RP REVISION TO 24.

```

```

RA Renshaw R.W.
RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF
CC INFECTION IS REQUIRED FOR VIRUS ASSEMBLY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; M55918; AAA91832.1; -
DR EMBL; M81223; AAA42882.1; -
DR EMBL; L14767; AAD09422.1; -
DR PIR; A39926; A39926.
DR PIR; A48343; A48343.
DR InterPro; IPR006945; Circo_VP2.
DR Pfam; PF04861; Circo_VP2; 1.
DR CONFLICT .153 153 V -> A (IN REF. 1).
FT CONFLICT 187 187 D -> N (IN REF. 2).
SQ SEQUENCE 216 AA; 24138 MW; F82B69EF88BDE52 CRC64;

Query Match
Best Local Similarity 5.8%; Score 95.5; DB 1; Length 216;
Matches 54; Conservative 44; Mismatches 74; Indels 103; Gaps 14;

QY 17 GGPAPAGAAKMKVVEPNAFGVNPFPLPQASRLQAKRDPSPVSGPVHLFRLSGKCFSLYE 76
DB 21 GQPGPSGAAGGVYISN-----ERSR-----RSTRITNGVQ 52
QY 77 STYKFECPFHNTQHEQTFRNAYSGILGIW-----HEWEIAN-NFTFGMMNRDGDAC 129
DB 53 ATKKFTAVGNPSIQRPDPWYRN-VNHSIAVWLRECSRHAKICNGQFRKGMFQECAGL 111
QY 130 RRSRSRQSKVELACGSKNRLA-----HYSESTCYALTFTFPLVCHPHALLVYPT 179
DB 112 EDRSTQASLEBALRPLRVQGRKARKLDYHSQPT-----PNKKKYVKT 156
QY 180 LPEALQROMQVQDVLADDELITPQGEKILRTLFEDAGYLTKEPENEPTQLEGSPDSLGF 239
DB 157 V-----RW-----KDELAD-----READF--TPSER--DGGTTSDF 164
QY 240 ETLNCRKAHKELSKEIKLGLTQ-HGIPYTRP 273
DB 185 D-----EDINFDIGDGSIVDELLGRPFYTP 210

RESULT 6
VP2_CAV82
ID_VP2_CAV82 STANDARD; PRT; 216 AA.
AC PS4093;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Protein VP2.
OS Chicken anemia virus (Japanese isolate 82-2) (CAV).
OC Viruses; ssDNA viruses; Circoviridae; Gyrovirus.
OX NCBI_TaxID=73476;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95297149; PubMed=7778281;
RA Kato A., Fujino M., Nakamura T., Ishihara A., Otaki Y.;
RT "Gene organization of chicken anemia virus.";
RL Virology 209:480-488(1995).
CC -!- FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF
CC INFECTION IS REQUIRED FOR VIRUS ASSEMBLY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----

```

```

CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; D31965; BAA06732.1; -
DR InterPro; IPR006945; Circo_VP2.
DR Pfam; PF04861; Circo_VP2; 1.
SQ SEQUENCE 216 AA; 24138 MW; E23169F592BDDE52 CRC64;

Query Match
Best Local Similarity 5.8%; Score 94.5; DB 1; Length 216;
Matches 54; Conservative 44; Mismatches 74; Indels 103; Gaps 14;

QY 17 GGPAPAGAAKMKVVEPNAFGVNPFPLPQASRLQAKRDPSPVSGPVHLFRLSGKCFSLYE 76
DB 21 GQPGPSGAAGGVYISN-----ERSR-----RSTRITNGVQ 52
QY 77 STYKFECPFHNTQHEQTFRNAYSGILGIW-----HEWEIAN-NFTFGMMNRDGDAC 129
DB 53 ATKKFTAVGNPSIQRPDPWYRN-VNHSIAVWLRECSRHAKICNGQFRKGMFQECAGL 111
QY 130 RRSRSRQSKVELACGSKNRLA-----HYSESTCYALTFTFPLVCHPHALLVYPT 179
DB 112 EDRSTQASLEBALRPLRVQGRKARKLDYHSQPT-----PNKKKYVKT 156
QY 180 LPEALQROMQVQDVLADDELITPQGEKILRTLFEDAGYLTKEPENEPTQLEGSPDSLGF 239
DB 157 V-----RW-----KDELAD-----READF--TPSER--DGGTTSDF 164
QY 240 ETLNCRKAHKELSKEIKLGLTQ-HGIPYTRP 273
DB 185 D-----EDINFDIGDGSIVDELLGRPFYTP 210

RESULT 7
VP2_CAV26
ID_VP2_CAV26 STANDARD; PRT; 216 AA.
AC PS4092;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Protein VP2.
OS Chicken anemia virus (USA isolate 26p4) (CAV).
OC Viruses; ssDNA viruses; Circoviridae; Gyrovirus.
OX NCBI_TaxID=73477;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341490; PubMed=1908516;
RA Claessens J.A.J., Schrier C.C., Mockett A.P.A., Jagt E.H.J.M.,
RA Sondermeijer P.J.A.;
RT "Molecular cloning and sequence analysis of the genome of chicken
RT anaemia agent.";
RL J. Gen. Virol. 72:2003-2006(1991).
CC -!- FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF
CC INFECTION IS REQUIRED FOR VIRUS ASSEMBLY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; D10068; BAA00959.1; -
DR InterPro; IPR006945; Circo_VP2.
DR Pfam; PF04861; Circo_VP2; 1.
SQ SEQUENCE 216 AA; 24066 MW; CD4B69EF88BDDC72 CRC64;

Query Match
Best Local Similarity 5.7%; Score 92.5; DB 1; Length 216;
Matches 54; Conservative 43; Mismatches 75; Indels 103; Gaps 14;

```

```

QY 17 GGPAPAGAAKMKVVEEPNAPFGVNNPFLPQASRLAKRDPSPVSGVHFLRSGKCSFLVE 76
DB 21 GQPGSGAAQGVIN-----ERSPR-----RISTTINGVQ 52
QY 77 STYKXFCFHHNVTOHEQTFRNNAVSGILGIW-----HEWEIAN-NTFTGMWRDADAC 129
DB 53 ATKNEPTAVGNPSLQADPDWYRW-NYHSLAVLRECSRSKAKI CNCGFRGHMFOECAGL 111
QY 130 RRSRSOSKVELACGSNRLA-----HVSESTCVALLTETPLVCHPHALLVYPT 179
DB 112 EDRSTQASLEELPLFLVQGRAPKRLDYHSQPT-----PNEKKVYKT 156
QY 180 LPEALQROMDQVEODLADLITPQGHKXLTLPEDAGYKLTPEENPTQLGGPDSIGF 239
DB 157 V-----RW-----QSLAD-----READF--TPSEB-----DGGTSSDF 184
QY 240 ETLNCRKAHKLSEIKRLKLTQ HGIPYTPR 273
DB 185 D-----GDINPDIGDSGLVDELLGRFFTP 210

RESULT 8
ID INVO_AOTTR STANDARD; PRT; 544 AA.
AC P24708;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Involutrin.
GN IVL.
OS Aotus trivirgatus (Night monkey) (Douroucouli).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotidae; Aotus.
OX NCBI_Taxid=9505;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90014142; PubMed=2507864;
RA Tseng H., Green H.;
RT "The Involutrin gene of the owl monkey: origin of the early region."
RL Mol. Biol. Evol. 6:460-468(1989).
CC -1- FUNCTION: Part of the insoluble cornified cell envelope (CE) of stratified squamous epithelia.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Constituent of the scaffolding of the cornified envelope.
CC -1- TISSUE SPECIFICITY: Keratinocytes of epidermis and other stratified squamous epithelia.
CC -1- PTM: Substrate of transglutaminase. Specific glutamines or lysines are cross-linked to keratins, desmoplakin and to inter involutrin molecules.
CC -1- SIMILARITY: Belongs to the involutrin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M25313; AAA5375.1; -
CC DR PIR; I36911; I36911.
CC DR InterPro; IPR002360; Involutrin.
CC DR InterPro; IPR000354; Involutrin_rpt.
CC DR Pfam; PF00904; Involutrin_39.
CC DR PROSITE; PS00795; INVOLUCRIN; 1.
CC KM Keratinization; Repeat.
CC SQ SEQUENCE 544 AA; 63927 MW; 2A02ABA5E1499F9D CRC64;

Query Match 5.6%; Score 92; DB 1; Length 544;
Best Local Similarity 23.4%; Pred. No. 4.2;
Matches 34; Conservative 25; Mismatches 48; Indels 38; Gaps 4;
QY 182 EALQROMDQVEODLADLITPQGHKXLTLPEDAGYKLTPEENPTQLGGPDSIGF 215
-----LITPQGHKXLTLPED 215

```

```

DB 110 EKAQREKQQLQGLQLEEKXKLLDQPDHKLAKSDQQLGTRKQQLTFEPQEQGLKCLBQ 169
QY 216 AGYLTPEENPTQLQEGGDSIGFTELNCRKAHKLSEIKR-RLKGLLTQHGITYPT 274
DB 170 EGHLEPEQEQEQ-----LKCLBQ-QEGHLEPEQEQGLKLEQEGKXKLEQ 218
QY 275 ETSNLEHLEHTPRAPSPPEQLRDP 239
DB 219 QEGQVHLEEQEKQSELPEQQRGP 243

RESULT 9
ID ML12_HUMAN STANDARD; PRT; 5262 AA.
AC 014686; 014687;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related protein).
GN ML12 OR ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX MEDLINE=97388474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bulirich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia and to Drosophila trithorax."
RL Oncogene 15:549-560(1997).
RN [2]
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=22371496; PubMed=12482968;
RA Goo Y.-H., Sohn Y.-C., Kim D.-H., Kim S.-N., Kang M.-J., Jung D.-J.,
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
RA Acores D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-D., Lee Y.C.,
RA Lee J.W.;
RT "Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins."
RL Mol. Cell. Biol. 23:140-149(2003).
CC -1- FUNCTION: May be involved in transcriptional regulation.
CC -1- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/ RBBP5, alpha- and beta-tubulins, the trithorax group proteins ML12 and ML3, and ASH2/ASCL2.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=O14686-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O14686-2; Sequence=VSP_008563, VSP_008559;
CC Name=3;
CC IsoId=O14686-3; Sequence=VSP_008560;
CC -1- TISSUE SPECIFICITY: Expressed in most adult tissues, including a variety of hematopoietic cells, with the exception of the liver.
CC -1- MISCELLANEOUS: This gene mapped to a chromosomal region involved in duplications and translocations associated with cancer.
CC -1- SIMILARITY: Belongs to the transcription factor trithorax family.
CC -1- SIMILARITY: Contains 5 PHD-type zinc fingers.
CC -1- SIMILARITY: Contains 1 post-SET domain.
CC -1- SIMILARITY: Contains 1 RING-type zinc domain.
CC -1- SIMILARITY: Contains 1 SET domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


[illegible]

DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Tetratricopeptide repeat protein 3 (TPR repeat protein D).
 GN TTP3 OR TTPD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96281123; PubMed=8724848;
 RA Chiba M., Ootsubayama A., Suzuki E., Ichikawa H., Seki N.,
 RA Nagase T., Nomura N., Ohki M.
 RT "Identification of a novel human gene containing the
 RT tetratricopeptide repeat domain from the Down syndrome region of
 RT chromosome 21."
 RL DNA Res. 3:9-16(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=97103476; PubMed=8947847;
 RA Tsukahara F., Hattori M., Muraki T., Sakaki Y.
 RT "Identification and cloning of a novel cDNA belonging to
 RT tetratricopeptide repeat gene family from Down syndrome-critical
 RT region 21q22.2."
 RL J. Biochem. 120:820-827(1996).
 RN [3]
 RP SEQUENCE OF 121-616 FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=98163741; PubMed=9503011;
 RA Dahmane N., Alt Ghezala G., Gossel P., Chamoun Z.,
 RA Dufresne-Zacharia M.C., Lopes C., Rabatel N., Gassanova-Maugendre S.,
 RA Chetoui Z., Abramowski V., Fayet E., Yaspo M.L., Korn B.,
 RA Blouin J.L., Leherich H., Pousterka A., Antonarakis S.E., Sinet P.M.,
 RA Creau N., Delabar J.M.
 RT "Transcriptional map of the 2.5-Mb CBR-ERG region of chromosome 21
 RT involved in Down syndrome."
 RL Genomics 48:12-23(1998).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=TRPD1;
 CC IsoId=PS3804-1; Sequence=Displayed;
 CC Name=TRPD1;
 CC IsoId=PS3804-2; Sequence=VSP_006554;
 CC Name=TRPD1;
 CC IsoId=PS3804-3; Sequence=VSP_006555;
 CC -1- TISSUE SPECIFICITY: Found in all tissues examined.
 CC -1- SIMILARITY: Contains 4 TPR repeats.
 CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D83077; BAA11769.1; -
 CC EMBL; D84294; BAA12301.1; -
 CC EMBL; D84295; BAA12302.1; -
 CC EMBL; D84296; BAA12303.1; -
 CC EMBL; A7001866; CAA05057.1; -
 CC PIR; JCS020; JCS020
 CC Gene; HGNC:12593; TTC3.
 CC WIM; 602259; -
 CC GO; GO:0009405; P:apathogenesis; TAS.
 CC InterPro; IPR008940; Ptenyl_trans.
 CC InterPro; IPR001440; TPR.
 CC InterPro; IPR001841; Znf_ring.
 CC Pfam; PF00515; TPR; 3.

DR Pfam; PF00097; zf-C3HC4; 1.
 DR SMART; SM00184; RING; 1.
 DR SMART; SM00028; TPR; 3.
 DR PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
 DR PROSITE; PS00089; ZF_RING_2; 1.
 DR PROSITE; PS00518; ZF_RING_2; 1.
 KW Repeat; TPR repeat; Zinc-finger; Alternative splicing; Polymorphism.
 FT REPEAT 231 264
 FT REPEAT 266 298
 FT REPEAT 366 572
 FT REPEAT 576 609
 FT ZN_RING 1957 1997
 FT DOMAIN 453 456
 FT DOMAIN 1020 1029
 FT DOMAIN 1899 1902
 FT DOMAIN 1018 1029
 FT DOMAIN 1172 1185
 FT DOMAIN 1563 1579
 FT VARSPLIC 1 233
 FT VARSPLIC 1 310
 FT VARIANT 1117 1117
 FT CONFLICT 121 121
 FT CONFLICT 139 139
 FT CONFLICT 232 232
 FT CONFLICT 276 276
 SQ SEQUENCE 2025 AA; 229889 MW; 1B4BCAA3684B6253 CRC64;
 Query Match 5.4%; Score 89; DB 1; Length 2025;
 Best local Similarity 21.5%; Pred. No. 41;
 Matches 65; Conservative 46; Mismatches 123; Indels 68; Gaps 14;
 QY 26 KKKVEEPNAPGVNPPFLQASRLQ--AKRDP--PVSG-VHLFRLSKGCEFLVST 78
 DB 1325 REPDTQTPPAYINVAIPGLPYQYISYIPLASLSPYQLPSPVSVPSFVANDADKNAAY 1384
 QY 79 YKYEPCPFNNVQHEQTFPMNAVSGILGIMHEMEJANNFTMMNRDGDACRSRQSKY 138
 DB 1385 FEGLHNAENVAGHQJASTQLLESLG---SVKSHST-----GDHVLVSESNRN 1434
 QY 139 ELAAGSKNRLAHVSESTCVVALTFETPIVCPHALLVPTLPALQROMQOVQDLADE 198
 DB 1435 DEHGNSNNKCEVTPST-----SAVTNIPHQMV-----AIQVSWNIHQEVNTE 1480
 QY 199 LITP-----GG-----HEKL---LRTLEAGVILKTP-----EENB- 226
 DB 1481 PYNPEERQGEISRIKKEHQVQDOQVYENYEQIKLGLBETRDLEKLRHLBENKI 1540
 QY 227 -PTLEGSDPSLGFETLENCRAKELSKIRLKL---TQHGIPYTRPTETSNLEHL 282
 DB 1541 SKTELDWPIQDLERE-IKKQGEKKEIQERLSLKKKIKVSNASBMYTQKNDGKEKEHE 1599
 QY 283 GH 284
 DB 1600 LH 1601
 RESULT 11
 STB2_CANFA
 ID STB2_CANFA STANDARD; PRT; 593 AA.
 AC Q28288;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Syntaxin binding protein 2 (Unc-18 homolog 2) (Unc-18B) (Unc18-2).
 GN STXB2
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 NC NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.

RL Nature 387:90-93(1997).
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z48430; CAA88330.1; -
 DR PIR; S52481; S52481.
 DR SGD; S0004503; YML039W.
 DR InterPro; IPR001969; Asparticase_AS.
 DR InterPro; IPR009007; Pept_A_acid.
 DR InterPro; IPR001584; Rve.
 DR Pfam; PF00665; rve; 1.
 DR PROSITE; PS00141; ASP_PROTEASE; 1.
 KW Transposable element; Hypothetical protein; Hydrolase;
 KW Aspartyl protease; ATP-binding
 FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
 FT NP_BIND 1204 1211 ATP (POTENTIAL).
 SQ SEQUENCE 1328 AA; 151036 MW; C7D14E1A4A675E93 CRC64;
 Query Match 5.4%; Score 88; DB 1; Length 1328;
 Best Local Similarity 22.9%; Pred. No. 29;
 Matches 77; Conservative 34; Mismatches 85; Indels 140; Gaps 22;
 QY 55 PEPVSGP---VHLRLSGKCSLVESTYKTEFCFENVTOG--EOTFRNAYGIGLIGM 108
 DB 146 PENISVPTNNVH-----TSESTRKYVPFIHMLHANNQITRYS----- 186
 QY 109 HEMELIANFTGMWRDGDACRSRSGKVELAC-----GKSNRLAHV----- 151
 DB 187 ---LKNNTIT--YNEBSVDNRS---SAIDQCDCIGSTKRHKSGRLKQNSYE 236
 QY 152 -----SEPTCYVALTFET-----PLVCHPH---ALLVPTLPE 182
 DB 237 PFOYLHTDIFGPVHNLPKSPSYFI-SFTDETKFRWVPL--HNRREDSTILDVFTTILA 293
 QY 183 ALQORWDQ---VEODLDEL-----ITP-----OGH---EKLRLTFE 214
 DB 294 FIKNGQASVLVIQMDRGEYTNRLHKLFLKNGITPCYTTADSRAHGAERLNLILD 353
 QY 215 DAGYKTPRENEPTOLE--GGPDSLGFTLENCRAKHEL-----SKEIKR---LKGLLT 264
 DB 354 DC-----RTQDCSGLEPHLMFSAIEPSTIVRNSLASPKSKGAPQAGIAGLDI 403
 QY 265 QHGIPYTRPTETSNLEHLGHTPRAKSPBQLRGDPG 300
 DB 404 STLLPFGQPIVND-----HNPNSKI--HPRGIPG 431
 RESULT 14
 Y9B8_PSEAR STANDARD; PRT; 364 AA.
 AC P33642; Q51527; Q9HVN1;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable D-amino acid oxidase PA4548.
 GN PA4548.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RX MEDLINE=96272255; PubMed=8682785;
 RA Alm R.A.; Mattick J.S.;
 RT Identification of two genes with prepilin-like leader sequences

RT Involved in type 4 fimbrial biogenesis in Pseudomonas aeruginosa.";
 RL J. Bacteriol. 178:3809-3817(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Strover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Matman S., Yuan X.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lavis K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RT Nature 406:959-964(2000).
 RL Nature 406:959-964(2000).
 RN [3]
 RP SEQUENCE OF 193-364 FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=9325810; PubMed=8097014;
 RA Hobbs M., Collier E.S.R., Free P.D., Livingston S.P., Mattick J.S.;
 RA Pils and PilR, a two-component transcriptional regulatory system
 RT controlling expression of type 4 fimbriae in Pseudomonas
 RT aeruginosa.";
 RL Mol. Microbiol. 7:669-682(1993).
 CC -1- COFACTOR: FAD (Potential).
 CC -1- SIMILARITY: Belongs to the dddA family of oxidoreductases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; I48934; AAB39269.1; -
 DR EMBL; AE004668; AAG07936.1; -
 DR EMBL; Z12154; CAA78142.1; -
 DR PIR; B83078; B83078.
 DR InterPro; IPR006076; Fad_oxred.
 DR InterPro; IPR000205; NAD_BS.
 DR InterPro; IPR000594; Thif_domain.
 DR Pfam; PF01266; DAO; 1.
 KW Hypothetical protein; Oxidoreductase; Flavoprotein; FAD;
 KW Complete proteome.
 FT NP_BIND 5 19 FAD (ADP PART) (POTENTIAL).
 FT CONFLICT 20 20 E -> K (IN REF. 1).
 FT CONFLICT 23 23 L -> I (IN REF. 1).
 FT CONFLICT 57 57 S -> N (IN REF. 1).
 FT CONFLICT 109 109 R -> P (IN REF. 1).
 FT CONFLICT 193 245 IRGDKTLAAGAMSGELKPIGLIEPVYVKGOMTIYKCA
 FT DFLPRVTLAKGR -> DPKRKAAGGRLEKRYVERAWPMT
 FT ARGTERSDPQVKGGEFPAAGAGQGG (IN REF. 3).
 FT ASA -> VSV (IN REF. 1).
 FT CONFLICT 280 282
 SQ SEQUENCE 364 AA; 39445 MW; 86C82DC62AA7A811 CRC64;
 Query Match 5.4%; Score 87.5; DB 1; Length 364;
 Best Local Similarity 20.4%; Pred. No. 6;
 Matches 56; Conservative 28; Mismatches 77; Indels 113; Gaps 11;
 QY 23 GAAMKVEBPNFAFGVNNPFLPQASRLQAKRDPSPVSGFVHLRLSGKCSLVESTYKXE 82
 DB 131 GAGFGRAVVMGSAVYANRFLAKRSLSAQ----- 161
 QY 83 FCFPHVTOHEQTFRNAYSGILGIWHEIANTFTGMMDP-----ACRSRSPS 136
 DB 162 ---FANLELHEQF-----EVAG-----WLRDGRVGVATSRSEIKRD 196
 QY 137 KVELACGK-SNRLAVSEPTCYVALTFETPLVCHPHALLVPTPEALQRCMDVEDPDL 195
 DB 197 KTLAAGAMSGEL-----LKPLGLELPVVPVKGOMTIYKCAADFLPR----- 238
 QY 196 ADELITPQGH-----EKLRLTFEADAGYKTPRENEPTOLEGGPDSLGFTLENCRAK 247

```

Db      239 ---MVLAKRVAIPRRDHILGSLHSGFDTKD-----EAGESURA 280
Qy      248 AHKEUSKEKRLK-----GLL--TQHGIPYTRP 273
Db      281 SAAELPPELADWQPAWMAGLRPGSPGIPYIGP 314

RESULT 15
SYM_CHLITE
ID      SYM_CHLITE      STANDARD:      PRT:      703 AA.
AC      P59077;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DE      Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)
DE      (MeRS).
GN      METG OR METS OR CT0969.
OS      Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC      Chlorobium.
OX      NCBI_Taxid=1097;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=TLS / ATCC 49652 / DSM 12025;
RX      MEDLINE=22103685; PubMed=12093901;
RA      Eissen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA      Dodson R.J., Deboy R., Gwin M.L., Nelson W.C., Haft D.H.,
RA      Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA      Holt I., Unayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA      Nierman W.C., Feldblyum T.V., Hansen C.L., Craven W.B., Radune D.,
RA      Vamathevan J., Khouri H., Bryant D.A., Fraser C.M.,
RA      Venter J.C., Tetteh H., White O., Gruber T.M., Ketchum K.A.,
RT      "The complete genome sequence of Chlorobium tepidum TLS, a
RT      photosynthetic, anaerobic, green-sulfur bacterium.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
CC      -!- FUNCTION: Is required not only for elongation of protein synthesis
CC      but also for the initiation of all mRNA translation through
CC      initiator tRNA (Met) aminoacylation (By similarity).
CC      -!- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +
CC      diphosphate + L-methionyl-tRNA(Met).
CC      -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC      -!- SUBUNIT: Homodimer (By similarity).
CC      -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC      MetG subfamily 1.
CC      -!- SIMILARITY: Contains 1 tRNA-binding domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE012861; AAM72204.1; -.
DR      TIGR; CT0969; -.
DR      HAMAP; MF_00098; fused; 1.
DR      InterPro; IPR004495; MetG_Cterm.
DR      InterPro; IPR008224; MeRS dimerising.
DR      InterPro; IPR008994; Nucleic acid OB.
DR      InterPro; IPR002300; tRNA-synt_1a.
DR      InterPro; IPR001412; tRNA-synt_1.
DR      InterPro; IPR002304; tRNA-synt_mec.
DR      InterPro; IPR002547; tRNA bind.
DR      Pfam; PF00133; tRNA-synt_1; 1.
DR      Pfam; PF01588; tRNA bind. 1.
DR      PIRSF; PIRSF001528; MeRS dimerising; 1.
DR      PRINTS; PR01041; TRNASYNTHEM.
DR      TIGRFAMs; TIGR00399; metG; 1.
DR      TIGRFAMs; TIGR00399; metG_Cterm; 1.
DR      PROSITE; PS00178; AA_tRNA_LIGASE_I; 1.

```

```

DR      PROSITE; PS50886; TRBD; 1.
KM      Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KM      RNA-binding; tRNA-binding; Metal-binding; Zinc; Complete proteome.
FT      SITE      15      25      "HIGH" REGION.
FT      SITE      345      349      "KMSKS" REGION.
FT      DOMAIN      602      703      tRNA-BINDING.
FT      METAL      147      147      ZINC (BY SIMILARITY).
FT      METAL      150      150      ZINC (BY SIMILARITY).
FT      METAL      160      160      ZINC (BY SIMILARITY).
FT      METAL      163      163      ZINC (BY SIMILARITY).
FT      BINDING      348      348      ATP (BY SIMILARITY).
SQ      SEQUENCE      703 AA; 79553 MW; 9C4440919E7F52E6 CRC64;

Query Match      5.4%; Score 87.5; DB 1; Length 703;
Best Local Similarity 25.3%; Pred. No. 14;
Matches 40; Conservative 25; Mismatches 50; Indels 43; Gaps 9;

Qy      125 DGDACRSRSRQSKVELACGKSNRIAHVSEP-----STCVYALTFETPLVCHPH 172
Db      448 DGFHFREATQGT-MEIA-RFANNFLTSESEPMKVIKYDPEAGRTMAVSLN-----LCHTL 500
Qy      173 ALVYPTLPALQRODQVE-QDLADELITPOCHEKILITFEDAGYLTPEENPTQLE 231
Db      501 ALFWPIVPEPTANRIKMGFGSTIDLEVP-GNPVWRQAL-----EPLGK 546
Qy      232 G---GPDSLGFETLNCRAHKEISKETRLKGLTQ 265
Db      547 GHKLSSSEILFSEKID-----KDIPKMKITALLAE 579

```

Search completed: July 26, 2004, 11:09:46
 Job time : 7.08008 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: Jul 26, 2004, 11:05:03 ; Search time 19.9295 Seconds
(without alignments)
4828.666 Million cell updates/sec

Title: US-10-657-280-3
Perfect score: 1634
Sequence: 1 MAAGLARIILLILGLSAGSPA.....TPRAKSPQLRGDPGLRGLSL 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL.25.*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_ricent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriopl:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1634	100.0	305	4	Q9UJ39
2	1618.5	99.1	304	4	Q9EL13
3	1601	98.0	331	4	Q9ER22
4	1301.5	79.7	307	11	Q7PNE0
5	1288	78.8	314	11	Q8CSJ3
6	185.5	11.4	507	5	Q23440
7	172.5	10.6	529	13	Q80222
8	171	10.5	647	10	Q9FM96
9	171	10.5	647	10	Q8LDD0
10	155.5	9.5	612	4	Q95W99
11	155.5	9.5	615	4	Q81Z58
12	148	9.1	548	5	Q9VUD1
13	141	8.6	521	11	Q08795
14	141	8.6	528	11	Q921X2
15	140.5	8.6	200	4	Q96BU9
16	140.5	8.6	398	4	Q96D06

17	140.5	8.6	528	4	Q9P0W9	Q9P0W9 homo sapien
18	130	8.0	571	11	Q922S5	Q922S5 mus musculu
19	130	8.0	617	11	Q8K2C7	Q8K2C7 mus musculu
20	125	7.6	483	4	Q9NUY7	Q9NUY7 homo sapien
21	120.5	7.4	483	11	Q8VEH8	Q8VEH8 mus musculu
22	120.5	7.4	483	11	Q8BYN6	Q8BYN6 mus musculu
23	119	7.3	182	4	Q9UOL4	Q9UOL4 homo sapien
24	119	7.3	306	4	Q95901	Q95901 homo sapien
25	119	7.3	483	4	Q96D21	Q96D21 homo sapien
26	117	7.2	482	13	Q803G0	Q803G0 brachydanto
27	116.5	7.1	498	11	Q9CWA6	Q9CWA6 mus musculu
28	110.5	6.8	352	4	Q86WL7	Q86WL7 homo sapien
29	106.5	6.5	177	10	Q94AF1	Q94AF1 aradidopsis
30	106	6.5	150	10	Q65221	Q65221 aradidopsis
31	105.5	6.5	352	4	Q86VFI	Q86VFI homo sapien
32	105.5	6.5	377	4	Q86V48	Q86V48 homo sapien
33	102.5	6.3	525	5	Q9VKC7	Q9VKC7 drosophila
34	100	6.1	534	5	Q9NFR2	Q9NFR2 caenorhabd
35	98.5	6.0	216	12	Q91ZU7	Q91ZU7 chicken ane
36	96.5	5.9	216	12	Q91N82	Q91N82 chicken ane
37	95.5	5.8	216	12	Q9DXA8	Q9DXA8 chicken ane
38	95.5	5.8	216	12	Q8QY49	Q8QY49 chicken ane
39	95.5	5.8	216	12	Q95670	Q95670 chicken ane
40	95.5	5.8	216	12	Q91NA6	Q91NA6 chicken ane
41	95.5	5.8	282	10	Q8GWH3	Q8GWH3 aradidopsis
42	95.5	5.8	1839	2	Q30765	Q30765 streptomyces
43	95.5	5.8	4290	2	Q9WXC0	Q9WXC0 micromonospora
44	95.5	5.8	4307	2	Q83WFO	Q83WFO micromonospora
45	95	5.8	407	16	Q66578	Q66578 aquifex aeo

ALIGNMENTS

RESULT 1

Q9UJ39 PRELIMINARY; PRT; 305 AA.
ID Q9UJ39
AC Q9UJ39
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE C316G12.3 (CAB56184).
GN C316G12.3 OR CAB56184.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Frankland J.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Fitzgerald P., Amarante-Wendes G.P., Li W., Green D.R.;
RT "DNA from human fetal brain."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031709; CAB56184.1; -;
DR EMBL; AF02786; AAG27706.1; -;
DR PIR; T45062; T45062.
DR InterPro; IPR009011; Man_6_P_R_bind.
SQ SEQUENCE 305 AA; 33973 MW; 7774BBC091DA1C2 CRC64;

Query Match 100.0%; Score 1634; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.2e-148;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARIILLILGLSAGSPAAGAAKCKVEEENAFGVNPFILPQASRLQAKRDPSPVG 60
DB 1 MAAGLARIILLILGLSAGSPAAGAAKCKVEEENAFGVNPFILPQASRLQAKRDPSPVG 60
QY 61 PVHLFRSLGKCELVSTYKFERCPTHNTQHEQTRNNAISGILGIMHEWELNANTFFG 120

```

Db      61  PVLHRLSGKCSLYESTYKVFPCPHNVTOGHEQTFRNMAVSGILGIWHEMIANNFTG 120
Qy      121  MMARDGDACRSRSRQSVKELACGKSNRLAHVSEPTCYVALLFTETPLVCHPALLVPTL 180
Db      121  MMARDGDACRSRSRQSVKELACGKSNRLAHVSEPTCYVALLFTETPLVCHPALLVPTL 180
Qy      181  PEALOROMQOVQODLADDELITPOGHEKILRTLFEDAGYLTKEPENEPTOLGSGPDSIGFE 240
Db      181  PEALOROMQOVQODLADDELITPOGHEKILRTLFEDAGYLTKEPENEPTOLGSGPDSIGFE 240
Qy      241  TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPETSJLEHLGHETPRAKSPQOLRGDPG 300
Db      241  TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPETSJLEHLGHETPRAKSPQOLRGDPG 300
Qy      301  LRGS 305
Db      301  LRGS 305

RESULT 2
Q96L13  PRELIMINARY; PRT; 304 AA.
ID      096L13
AC      096L13,
DT      01-DEC-2001 (TREMBlrel. 19, Created)
DT      01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Hypothetical protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RA      Strausberg R.;
RL      Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC014592; AAK14592.1; -.
DR      InterPro; IPR009011; Man_6_P_R_bind.
KW      Hypothetical protein.
SQ      SEQUENCE 304 AA; 33844 MW; FCIEASA393DBB6C9 CRC64;

```

```

Query Match      99.1%; Score 1618.5; DB 4; Length 304;
Best Local Similarity 99.7%; Pred. No. 3.7e-147;
Matches 304; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

```

Qy      1  MAAGLARLLLLGLSAGGAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
Db      1  MAAGLARLLLLGLSAGGAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
Qy      61  PVLHRLSGKCSLYESTYKVFPCPHNVTOGHEQTFRNMAVSGILGIWHEMIANNFTG 120
Db      61  PVLHRLSGKCSLYESTYKVFPCPHNVTOGHEQTFRNMAVSGILGIWHEMIANNFTG 120
Qy      121  MMARDGDACRSRSRQSVKELACGKSNRLAHVSEPTCYVALLFTETPLVCHPALLVPTL 180
Db      121  MMARDGDACRSRSRQSVKELACGKSNRLAHVSEPTCYVALLFTETPLVCHPALLVPTL 180
Qy      181  PEALOROMQOVQODLADDELITPOGHEKILRTLFEDAGYLTKEPENEPTOLGSGPDSIGFE 240
Db      181  PEALOROMQOVQODLADDELITPOGHEKILRTLFEDAGYLTKEPENEPTOLGSGPDSIGFE 240
Qy      241  TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPETSJLEHLGHETPRAKSPQOLRGDPG 300
Db      241  TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPETSJLEHLGHETPRAKSPQOLRGDPG 300
Qy      301  LRGS 305
Db      301  LRGS 305

```

```

RESULT 3
Q96R22  PRELIMINARY; PRT; 331 AA.
ID      Q96R22

```

```

AC      Q96R22;
DT      01-DEC-2001 (TREMBlrel. 19, Created)
DT      01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Similar to protein kinase C substrate.
GN      RGD9.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=21096910; PubMed=1157797;
RA      Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA      Tyfarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA      Higgs D.R.;
RT      "Sequence, structure and pathology of the fully annotated terminal 2
RT      kb of the short arm of human chromosome 16";
RL      Hum. Mol. Genet. 10:339-352(2001).
DR      EMBL; AE006467; AAK61277.1; -.
DR      GO; GO:0016301; P.kinase activity; IEA.
DR      InterPro; IPR009011; Man_6_P_R_bind.
KW      Kinase.
SQ      SEQUENCE 331 AA; 36799 MW; 0D01BBE475645F2F CRC64;

```

```

Query Match      98.0%; Score 1601; DB 4; Length 331;
Best Local Similarity 92.1%; Pred. No. 2e-145;
Matches 305; Conservative 0; Mismatches 0; Indels 26; Gaps 2;

```

```

Qy      1  MAAGLARLLLLGLSAGGAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 59
Db      1  MAAGLARLLLLGLSAGGAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
Qy      60  -----GPVHRLSGKCSLYESTYKVFPCPHNVTOGHEQTFRNMAVSGI 104
Db      61  RFGAWTLGISVNSLPFVHRLSGKCSLYESTYKVFPCPHNVTOGHEQTFRNMAVSGI 120
Qy      105  LGIWHMEIANNFTGMMARDGDACRSRSRQSVKELACGKSNRLAHVSEPTCYVALLTPE 164
Db      121  LGIWHMEIANNFTGMMARDGDACRSRSRQSVKELACGKSNRLAHVSEPTCYVALLTPE 180
Qy      165  TPLVCHPALLVPTLPEALOROMQOVQODLADDELITPOGHEKILRTLFEDAGYLTKEP 224
Db      181  TPLVCHPALLVPTLPEALOROMQOVQODLADDELITPOGHEKILRTLFEDAGYLTKEP 240
Qy      225  NEPTOLGSGPDSIGFETLNCRAHKELSKEIKRLKGLLTQHGIPYTRP----- 274
Db      241  NEPTOLGSGPDSIGFETLNCRAHKELSKEIKRLKGLLTQHGIPYTRPGEDWPBGVLA 300
Qy      275  ETSNLEHLGHETPRAKSPQOLRGDPGRSL 305
Db      301  ETSNLEHLGHETPRAKSPQOLRGDPGRSL 331

```

```

RESULT 4
Q7TNE0  PRELIMINARY; PRT; 307 AA.
ID      Q7TNE0
AC      Q7TNE0;
DT      01-OCT-2003 (TREMBlrel. 25, Created)
DT      01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Hypothetical protein.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=FVB/N; TISSUE=salivary gland;
RX      MEDLINE=22388257; PubMed=12477932;
RA      Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shellen G.M., Schuler G.D.,
RA      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,

```


Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Ditchenko L., Marinova K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinci F., Prange C.,
RA Raha S.S., Loguella N.A., Peters G.J., Adamson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Buterfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

Query Match 79.7%; Score 1301.5; DB 11; Length 307;
Best Local Similarity 78.8%; Pred. No. 1.2e-116;
Matches 241; Conservative 25; Mismatches 39; Indels 1; Gaps 1;

QY 1 MAAGIARLLILGLSAGGPAPGAAMKVEEPNAGVNNPFLPQASRLQKRPSPVSG 60
DB 1 MAGRLAGFLMLGLASQGPAPGACGMKVEEPNTPGLNPLPQASRLQKRPSPVSG 60
QY 61 PVLFLRSLGKCFSLVSTYKFCPEPHNTQHEQFRMNAVSGLIGIHEWEIANNFTG 120
DB 61 PVLFLRSLGKCFSLVSTYKFCPEPHNTQHEQFRMNAVSGLIGIHEWEIANNFTG 120
QY 61 PVLFLRSLGKCFSLVSTYKFCPEPHNTQHEQFRMNAVSGLIGIHEWEIANNFTG 120
DB 61 PVLFLRSLGKCFSLVSTYKFCPEPHNTQHEQFRMNAVSGLIGIHEWEIANNFTG 120
QY 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPRLVPTL 180
DB 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPRLVPTL 180
QY 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPRLVPTL 180
DB 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPRLVPTL 180
QY 181 PEALGRQNDQVQODLADDELITPQGHKILRTLFEDAGYLKTPBEENPTQLGSGDSGFE 240
DB 181 PEALGRQNDQVQODLADDELITPQGHKILRTLFEDAGYLKTPBEENPTQLGSGDSGFE 240
QY 241 TLNCRKAHKEISKEIKLKGILLTOHGIPTREPTETSNLEHGHETP-RAKSPQLRNDP 299
DB 241 TLNCRKAHKEISKEIKLKGILLTOHGIPTREPTETSNLEHGHETP-RAKSPQLRNDP 299
QY 300 GLRGSU 305
DB 301 GLRGN 306

RESULT 5
Q8CSJ3 PRELIMINARY; PRT; 314 AA.
ID 08CSJ3
AC 08CSJ3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical extracytoplasmic domain of cation-dependent mannose
6-phosphate receptor structure containing protein.
GN 1300004G08RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=22354683; PubMed=1246851;

RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
Nature 420:563-573 (2002).
RL Nature 420:563-573 (2002).
DR EMBL; AK078230; BAC37183.1; -
DR MGD; MGI:1921404; 1300004G08RIK.
DR InterPro; IPR009011; Man_6_P_R_bind.
KW Hypothetical protein.
SQ SEQUENCE 314 AA; 34891 MW; 7C3424A15649D058 CRC64;

Query Match 78.8%; Score 1288; DB 11; Length 314;
Best Local Similarity 77.0%; Pred. No. 2.4e-115;
Matches 241; Conservative 25; Mismatches 39; Indels 8; Gaps 2;

QY 1 MAAGIARLLILGLSAGGPAPGAAMKVEEPNAGVNNPFLPQASRLQKRPSPVSG 60
DB 1 MAGRLAGFLMLGLASQGPAPGACGMKVEEPNTPGLNPLPQASRLQKRPSPVSG 60
QY 61 PVLFLRSLGKCFSLVSTYKFCPEPHNTQHEQFRMNAVSGLIGIHEWEIANNFTG 120
DB 61 PVLFLRSLGKCFSLVSTYKFCPEPHNTQHEQFRMNAVSGLIGIHEWEIANNFTG 120
QY 61 PVLFLRSLGKCFSLVSTYKFCPEPHNTQHEQFRMNAVSGLIGIHEWEIANNFTG 120
DB 61 PVLFLRSLGKCFSLVSTYKFCPEPHNTQHEQFRMNAVSGLIGIHEWEIANNFTG 120
QY 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPRLVPTL 173
DB 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPRLVPTL 180
QY 174 LLYVPTLPEALGRQNDQVQODLADDELITPQGHKILRTLFEDAGYLKTPBEENPTQLGSG 233
DB 181 LLYVPTLPEALGRQNDQVQODLADDELITPQGHKILRTLFEDAGYLKTPBEENPTQLGSG 240
QY 234 PDSLGFETLENCRKAHKEISKEIKLKGILLTOHGIPTREPTETSNLEHGHETP-RAKSP 292
DB 241 SKGLGELTLDNCRKAHKEISKEIKLKGILLTOHGIPTREPTETSNLEHGHETP-RAKSP 300
QY 293 EOLRSDPGLRGN 305
DB 301 EHLRSDPGLRGN 313

RESULT 6
Q23440 PRELIMINARY; PRT; 507 AA.
ID Q23440
AC Q23440; Q22519;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ZK1307.8 protein.
GN ZK1307.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peleiderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Berk M.;
RT Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z47356; CAAB7420.1; -
DR EMBL; Z47358; CAAB7420.1; JOINED.
DR EMBL; Z47358; CAAB7438.1; -
DR EMBL; Z47356; CAAB7438.1; JOINED.
DR PIR; T24944; T24944.
DR WormPep; ZK1307.8; CB15547.
DR GO; GO:0005509; Focalization binding; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR00086; EF-hand.
DR InterPro; IPR002172; LDU_receptor_A.
DR InterPro; IPR009011; Man_6_P_R_bind.
DR Pfam; PF00036; ehand; 1.
DR SMART; SM00192; LduA; 1.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 507 AA; 58062 MW; 2764EC56741227DF CRC64;

Query Match 11.4%; Score 185.5; DB 5; Length 507;
 Best Local Similarity 38.7%; Pred. No. 5.5e-09;
 Matches 41; Conservative 11; Mismatches 47; Indels 7; Gaps 4;

QY 67 LSGKCFSLVSTYKTEFCFPHNVTOHEQTFRRMNAVSGI-LGIWHEWE-IANNFTFGMMMR 124
 DB 394 LNDKCGDRNVQYTYQFCFPGONTQKDT----GATSGTSLSGFKMSPEGNKYSKHAFG 449

QY 125 DGDAC-RSRSRQSKVELACGKSNRLAHVSEPTCYALTFETPLVC 169
 DB 450 DQOQCNGPKRSTDTITIEGGENELVEVTEPAKCEYLFTFTPLAC 495

RESULT 7
 Q80222 PRELIMINARY; PRT; 529 AA.

AC Q80222;
 DT 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DE Similar to protein kinase C substrate 80X-H.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NC NCB1_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC046883; AAH46883.1; .
 DR GO; GO:0005509; Ficolin ion binding; IEA.
 DR GO; GO:0016301; P-kinase activity; IEA.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR009011; Man 6_P_R_bind.
 DR PROSITE; PS00018; EF_HAND; 1.
 KM kinase.
 SQ SEQUENCE 529 AA; 59584 MW; BIEAF134CEE92502 CRC64;

Query Match 10.6%; Score 172.5; DB 13; Length 529;
 Best Local Similarity 36.2%; Pred. No. 1e-07;
 Matches 38; Conservative 12; Mismatches 50; Indels 5; Gaps 3;

QY 67 LSGKCFSLVSTYKTEFCFPHNVTOHEQTFRRMNAVSGI-LGIWHEWE-IANNFTFGMMMR 125
 DB 412 LYSQCYELSTSEYIYRLCPFNVSQKP---KFGSETVLGTGWSMGSPENNKYLSMKYEH 468

QY 126 GDAC-RSRSRQSKVELACGKSNRLAHVSEPTCYALTFETPLVC 169
 DB 469 GTGCMGPNRSTTVTLTGCKETMLTSTSEPSRCETLMEFTTPAVC 513

RESULT 8
 Q9FM96 PRELIMINARY; PRT; 647 AA.

AC Q9FM96;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Genomic DNA, chromosome 5, P1 clone:MCD7.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCB1_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=98290546; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.,
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.

RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned PL and TAC clones."
 RL DNA Res. 5:41-54(1998).
 DR EMBL; AB009049; BAB11263.1; .
 DR InterPro; IPR000886; ER target S.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR SMART; SM00192; LDLA; 1.
 DR PROSITE; PS00014; ER_TARGET; 1.
 SQ SEQUENCE 647 AA; 73213 MW; 49CC4FED7026472 CRC64;

Query Match 10.5%; Score 171; DB 10; Length 647;
 Best Local Similarity 28.0%; Pred. No. 1.9e-07;
 Matches 54; Conservative 24; Mismatches 77; Indels 38; Gaps 8;

QY 30 VEEPNAGVNNPFLPQASRL-----QAKRDPSPVSGPHLFRLSKCSLYES 77
 DB 479 VDKSEADRYRKEDSSSLNKISRLSEKLKQDFEPK--EFSFHCRCFESKQG 535

QY 78 TYKTEFCFPHNVTOHEQTFRRMNAVSGI-LGIWHEWE-IANNFTFGMMMRDGDAC-RSRSRQ 135
 DB 536 KYTYKVCAYKATQEE-----GYKTRLGEVDKE--NSYQFMSYNGKCKMGNPDRS 586

QY 136 SKVELACGKSNRLAHVSEPTCYALTFETPLVCPHALVPTLPEALQROMDOVEODL 195
 DB 587 LKYLKRCGLKRLMDVDEPSRCEYALISTPARC-----LEDKLELQQLKLEKLMND- 639

QY 196 ADELITPOGEXL 208
 DB 640 -----KPNHDEL 647

RESULT 9
 Q8LDD0 PRELIMINARY; PRT; 647 AA.

AC Q8LDD0;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCB1_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Haas B.D., Volfovsky N., Town C.D., Troupkan M., Alexandrov N.,
 RA Feldman K.A., Flavell R.B., White O., Salzberg S.L.;
 RT "Full-length messenger RNA sequences greatly improve genome
 annotation."
 RL Genome Biol. 0:0-0(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troupkan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldman K.;
 RT "Full-length cDNA from Arabidopsis thaliana."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AT086076; AAH63282.1; .
 DR InterPro; IPR000886; ER target S.
 DR PROSITE; PS00014; ER_TARGET; 1.
 KM Hypothetical protein.
 SQ SEQUENCE 647 AA; 73200 MW; 49C948C52P46B62 CRC64;

Query Match 10.5%; Score 171; DB 10; Length 647;
 Best Local Similarity 28.0%; Pred. No. 1.9e-07;
 Matches 54; Conservative 24; Mismatches 77; Indels 38; Gaps 8;

QY 30 VEEPNAGVNNPFLPQASRL-----QAKRDPSPVSGPHLFRLSKCSLYES 77
 DB 479 VDKSEADRYRKEDSSSLNKISRLSEKLKQDFEPK--EFSFHCRCFESKQG 535

QY 78 TYKTEFCFPHNVTOHEQTFRRMNAVSGI-LGIWHEWE-IANNFTFGMMMRDGDAC-RSRSRQ 135
 DB 536 KYTYKVCAYKATQEE-----GYKTRLGEVDKE--NSYQFMSYNGKCKMGNPDRS 586

QY 136 SKVELACGKSNRLAHVSEPTCYALTFETPLVCPHALVPTLPEALQROMDOVEODL 195
 DB 587 LKYLKRCGLKRLMDVDEPSRCEYALISTPARC-----LEDKLELQQLKLEKLMND- 639

QY 196 ADELITPOGEXL 208
 DB 640 -----KPNHDEL 647

Db 536 KATYVCAVKAATQBE-----GYSKTRLSEMDKFE--NSQFMSYTNKSGCANNPDRS 586
QY 136 SKVELACGSKSNRLAHVSEPTCYVALTFETPLVCHPHALVYPTLPALQROWDOVEODL 195
Db 587 LKVKLRGGLKNEIMVDDEPSRCEVAALISTPARC-----LEDXKLEIQQLERLXNMD- 639
QY 196 ADELITPOGHEKL 208
Db 640 -----KQONHDEL 647

RESULT 10
Q9BM99
ID Q9BM99 PRELIMINARY; PRT; 612 AA.
AC Q9BM99, 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to amplified in osteosarcoma.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC000532; AAH00532.1; -
DR InterPro; IPR009011; Man_6_P_R_bind.
SQ SEQUENCE 612 AA; 69246 MW; 1FD06F7B23101015 CRC64;

Query Match 9.5%; Score 155.5; DB 4; Length 612;
Best Local Similarity 21.7%; Pred. No. 5.5e-06;
Matches 79; Conservative 47; Mismatches 121; Indels 117; Gaps 18;

QY 21 PAGAAMKVVEBNFAGVNNPFLPQASRLQAKRDPSPVSGPVHLFRLSGKCFSLVESTYK 80
Db 76 PAGAHLFQREBERETPAVQGPGLPEL--LSPMRD-----APCLTKTDMWT 119
QY 81 YEFCEPHNVYQ-----HEQTFRNNAVSGILGIWHEWEIANNFTGMM 123
Db 120 YEFCEGRHIQGYHMEDESEIKGEVLYGYQSAFMDDEFTAKASKQRLKRYHSQTYG--- 176
QY 124 RQGDAC--RSRSQSVLELACGK-----SNRLAHVSEPTCYVALTFETPLVC----- 169
Db 177 -NGSKCDLNGRPREAEVRFLODEGAGISGDYIDRVDEPLSCSVLITIRPRLCPHPLRP 235
QY 170 ---HPHALVYPTL-PEA---LQROWDOVE-----ODLADEL-----ITPO- 203
Db 236 PPSAAPQALICHSLOPEBYMAVYORQADSKQYGDKIIEELQDLGQVWSETKSGVAPQK 295
QY 204 -----GHEKILRTLFEDA-GYLKTP-EENEPT-----QLEGPDLSGFETL 242
Db 296 MAGASPTKDDSKDQSPWKLNEPEQAPGGEVPAEEDPSPEADASGAPVDFQNNVQ 355
QY 243 ENCRKAHKEISKIKRLKGLTQHGIPYTRPETSLSLHGLHETP-----RAKSPEDL 295
Db 356 VAVISPADLIRFIELKG-GTKKGP-----NIGQEQVDDAAEVPOREPEKE 403
QY 296 RQDP 299
Db 404 RQDP 407

RESULT 11
Q81Z58
ID Q81Z58 PRELIMINARY; PRT; 615 AA.
AC Q81Z58, 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (Feb-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC023513; AAH23513.1; -
DR InterPro; IPR009011; Man_6_P_R_bind.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 615 AA; 69641 MW; 757D625B6FDA7097 CRC64;

Query Match 9.5%; Score 155.5; DB 4; Length 615;
Best Local Similarity 21.7%; Pred. No. 5.5e-06;
Matches 79; Conservative 47; Mismatches 121; Indels 117; Gaps 18;

QY 21 PAGAAMKVVEBNFAGVNNPFLPQASRLQAKRDPSPVSGPVHLFRLSGKCFSLVESTYK 80
Db 79 PAGAHLFQREBERETPAVQGPGLPEL--LSPMRD-----APCLTKTDMWT 122
QY 81 YEFCEPHNVYQ-----HEQTFRNNAVSGILGIWHEWEIANNFTGMM 123
Db 123 YEFCEGRHIQGYHMEDESEIKGEVLYGYQSAFMDDEFTAKASKQRLKRYHSQTYG--- 179
QY 124 RQGDAC--RSRSQSVLELACGK-----SNRLAHVSEPTCYVALTFETPLVC----- 169
Db 180 -NGSKCDLNGRPREAEVRFLODEGAGISGDYIDRVDEPLSCSVLITIRPRLCPHPLRP 238
QY 170 ---HPHALVYPTL-PEA---LQROWDOVE-----ODLADEL-----ITPO- 203
Db 239 PPSAAPQALICHSLOPEBYMAVYORQADSKQYGDKIIEELQDLGQVWSETKSGVAPQK 298
QY 204 -----GHEKILRTLFEDA-GYLKTP-EENEPT-----QLEGPDLSGFETL 242
Db 299 MAGASPTKDDSKDQSPWKLNEPEQAPGGEVPAEEDPSPEADASGAPVDFQNNVQ 358
QY 243 ENCRKAHKEISKIKRLKGLTQHGIPYTRPETSLSLHGLHETP-----RAKSPEDL 295
Db 359 VAVISPADLIRFIELKG-GTKKGP-----NIGQEQVDDAAEVPOREPEKE 406
QY 296 RQDP 299
Db 407 RQDP 410

RESULT 12
Q9VUD1
ID Q9VUD1 PRELIMINARY; PRT; 548 AA.
AC Q9VUD1, 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CG6453 protein (LD4653P).
GN CG6453.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Cainlaker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Ditzel S.M.,
 RA Dodson K., Doup L.E., Domene M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman I.U., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.W., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouenavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celinker S.,
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AE003655; AAF53621.1; -
 DR EMBL; AY058725; AAL13954.1; -
 DR FLYbase; FBgn0032643; CG6453.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000886; ER_target_S.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR009011; Man_6_P_bind.
 DR Pfam; PF00036; ehand; 1.
 DR SMART; SM00192; LDLA; 1.
 DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS00014; ER_TARGET; 1.
 DR PROSITE; PS50068; LDLA_2; 1.
 DR PROSITE; PS50068; LDLA_2; 1.
 SQ SEQUENCE 548 AA; 61539 MW; 4F486B724D64732E CRC64;

Query Match 9.1%; Score 148; DB 5; Length 548;
 Best Local Similarity 27.3%; Pred. No. 2.5e-05;
 Matches 36; Conservative 20; Mismatches 58; Indels 18; Gaps 4;

DB 434 GQCYNFEDREYVYTLCPFDRAQKS---RSGPPTTLGRWMDKSGEYQYKQYTGAA 490
 QY 69 GKCEVLVSTYKYECPFNHTYQHQTFRMAYSGILGTFEMIANNTFTGQWMDGDA 128
 DB 129 C-RSRRSQSKVELACGKSNRLAHVSEPTCYVALTFETPCVHPHLLVPTLPEALQRO 167
 DB 491 CMNGGNRAAIINISALPEKKTAVSEPRCEYFEFFETPAACD-----SEALQSE 540
 QY 188 WDQVEQDLDEL 199
 DB 541 ----SENLDDEL 548

RESULT 13
 ID 008795 PRELIMINARY; PRT; 521 AA.
 AC 008795
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Alpha glucosidase II, beta subunit.
 GN PRKCSH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAKRTLS 12.1/EL4;
 RX MEDLINE=97294720; Pubmed=9148925;
 RA Arendt C.W., Ostergaard H.L.,
 RT "Identification of the CD45-associated 116-kDa and 80-kDa proteins as
 the alpha- and beta-subunits of alpha-glucosidase II,"
 RL J. Biol. Chem. 272:13117-13125(1997).
 DR EMBL; U92794; AAC5183.1; -
 DR MGD; MGI:107877; PrkcsH.
 DR GO; GO:0017177; C:alpha-glucosidase complex; IPI.
 DR GO; GO:004558; F:alpha-glucosidase activity; IPI.
 DR GO; GO:000515; F:protein binding; IPI.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000886; ER_target_S.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR009011; Man_6_P_bind.
 DR Pfam; PF00036; ehand; 2.
 DR SMART; SM00192; LDLA; 1.
 DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS00014; ER_TARGET; 1.
 SQ SEQUENCE 521 AA; 58792 MW; BD070319898B4A38 CRC64;

Query Match 8.6%; Score 141; DB 11; Length 521;
 Best Local Similarity 31.0%; Pred. No. 0.00011;
 Matches 36; Conservative 15; Mismatches 35; Indels 12; Gaps 5;

DB 396 GSGEPAVLYVSCYELTWEYVYRLCPFLVSGKPK-----HGSPSTLGTWGSAGPD 449
 QY 60 GPVHLFR-LSGCFSLVSTYKYECPFNHTYQHQTFRMAYSG---ILGIHWE-IA 114
 DB 396 GSGEPAVLYVSCYELTWEYVYRLCPFLVSGKPK-----HGSPSTLGTWGSAGPD 449
 QY 115 NNTFTGMMWRDDAC-RSRRSQSKVELACGKSNRLAHVSEPTCYVALTFETPLVC 169
 DB 450 HDKFSAMKYEQGTGCGQGNRSTVRLGCKETVVTSTTSPRCYELMELMTPAAC 505

RESULT 14
 ID 0921X2 PRELIMINARY; PRT; 528 AA.
 AC 0921X2
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to protein kinase C substrate 80K-H.
 GN PRKCSH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC009816; AA09816.1; -
 DR MGD; MGI:107877; PrkcsH.
 DR GO; GO:0017177; C:alpha-glucosidase complex; IPI.
 DR GO; GO:004558; F:alpha-glucosidase activity; IPI.
 DR GO; GO:000515; F:protein binding; IPI.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000886; ER_target_S.

DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR009011; Man_6_P_R_bind.
 DR Pfam; PF00036; eHand; 2.
 DR SMART; SM00192; LDLa; 1.
 DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS00014; ER_TARGET; 1.
 SQ SEQUENCE 528 AA; 59559 MW; 253FD1A76A641713 CRC64;

Query Match 8.6%; Score 141; DB 11; Length 528;
 Best Local Similarity 31.0%; Pred. No. 0.00011;
 Matches 36; Conservative 15; Mismatches 53; Indels 12; Gaps 5;

QY 60 GPVHLFR-LSGKCFSLVESTKYKFCPEFHNVTQHEQTFRNAYSG--ILGIHWE-IA 114
 DB 403 GPSGFEALYSQCVELTNEVYRLCPFKLVSOXPK-----HGSPISLGTGSMNAGPD 456
 QY 115 NNTFTGMMWRDGDAC-RSRSRQSKVELACGKSNRLAHVSEPSCTCVYALTFETPLVC 169
 DB 457 HDKFSAMKYEOGTGCGPNRSTVRLCGKETVVTSTPEPRCEYLMELMTPAAC 512

RESULT 15

ID Q96BU9 PRELIMINARY; PRT; 200 AA.
 AC Q96BU9;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
 DE Hypothetical protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC015154; AAH15154.1; -.
 DR InterPro; IPR000886; ER_target.S.
 DR InterPro; IPR009011; Man_6_P_R_bind.
 DR PROSITE; PS00014; ER_TARGET; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 200 AA; 22628 MW; B9FA34051267DE9E CRC64;

Query Match 8.6%; Score 140.5; DB 4; Length 200;
 Best Local Similarity 30.5%; Pred. No. 3.3e-05;
 Matches 32; Conservative 14; Mismatches 54; Indels 5; Gaps 3;

QY 67 LSGKCFSLVESTKYKFCPEFHNVTQHEQTFRNAYSGILGIHWE-RIANNFTGMMWRD 125
 DB 83 LYSQCYELTNEVYRLCPFKLVSOXPK--KLGSPTSILGTGSMIGPDHDKFSAMKYEQ 139
 QY 126 GDAC-RSRSRQSKVELACGKSNRLAHVSEPSCTCVYALTFETPLVC 169
 DB 140 GTGCWQGPNRSTVRLCGKETVVTSTPEPRCEYLMELMTPAAC 184

Search completed: July 26, 2004, 11:11:40
 Job time : 22.9295 secs

This Page Blank (uspto)